


```
seq_name: /cgn2-6/ptdata/a/2/104/102105_0000 seq_101-0294-05412-1
seq_documentation_block:
: Sequence 1, Application PC/TUS9305412
: GENERAL INFORMATION:
: APPLICANT: Hirschberg, Carlos B.
: APPLICANT: Orellana, Ariel
: APPLICANT: Hashimoto, Yasuhiro
: APPLICANT: Swedler, Stuart J.
: APPLICANT: Wei, Zheng
: APPLICANT: Ishihara, Masayuki
```



```

202 yThrLysCysPheTyrCysLeu 209
|||||
3459 CCACCGGAGGAAATATGCGTC 3437

seq_name: /cn2_6/ptodata/2/ina/5a_311 x US-09-362-473-5
seq_documentation_block:
: Sequence 5, Application US/09362473
: Patent No. 6218169
: GENERAL INFORMATION:
: APPLICANT: Cahoon, Edgar B.
: APPLICANT: Cahoon, Rebecca E.
: APPLICANT: Falco, S. Carl
: APPLICANT: Morgante, Michele
: APPLICANT: Rafalski, J. Antoni
: APPLICANT: Hitz, William D.
: APPLICANT: Kinney, Anthony J.
: TITLE OF INVENTION: Aromatic Amino Acid Catalolism Enzymes
: FILE REFERENCE: BB-1197
: CURRENT APPLICATION NUMBER: US/09/362.473
: EARLIER APPLICATION NUMBER: 60/094,793
: EARLIER FILING DATE: JULY 31, 1998
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: Microsoft Office 97
: SEQ ID NO 5
: LENGTH: 1639
: TYPE: DNA
: ORGANISM: Glycine max
US-09-362-473-5

alignment_scores:
Quality: 87.50 Length: 213
Ratio: 0.833 Gaps: 7
Percent Similarity: 49.296 Percent Identity: 21.596

alignment_block:
US-09-557-262-2_copy_53_311 x US-09-362-473-5/rev ..
Align seq 1/1 to reverse of US 09 362 473-5 from 1 to 1639

61 PheSerProHisGlnLeuThrValGluLysThrProAlaTyrPheTh 77
|||||
604 TTCAACCTTCACATTCAGTAGACACAGGAGCTTCCTTCTGGGCAAC 555
|||||
77 rSerProLysValProGluArqLleHisSerMetAsnProThrIleArqL 94
|||||
554 TATCAGAGATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 505
|||||
94 cGluLeuLleLeuArqAspProSerGluArqValLeuSerAspTyrThr 110
|||||
504 TGTGGAGTGC .. TACATG 488

111 GlnValLeuTyrAsnHisLeuGlnLysHisLysProTyrProIleG 127
|||||
487 TCAATAGATAGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 438
|||||
127 uAspLeuLeuMetArqAspGlyArqLeuAsnLeuAspTyrLysAlaLeu 144
|||||
447 GAGC .. AATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 412

144 sArqSer .. LeuTyrHis .. 349
|||||
411 GCAATATGAGGATATCAATCAATCAATCAATCAATCAATCAATCAAT 362
|||||
150 .. .. AlaHisMetLeuAsnTrpLeuArqPhePheProLeuGlyHisI 164
|||||
361 CAACGTCGAGATGCTTAACTACATCACTCACTCACTCACTCACTCACT 312
|||||
164 eHisLeuValAspGly .. AspArqLeuLeuLeuArqAspProP 177
|||||

```

```

311 CCTAGGCTTTAAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 262
177 heProGluLleGlnLysValGluArgPheLeuLysLeuSerProGlnIle 193
|||||
261 AACTCAAGAGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 212
|||||
194 AsnAlaSerAsnPheTyrPheAsnLysThrLysGlyPheTyrCysLeuAr 210
|||||
211 TCTTCGGCGGTAGAGCGGCTAGCGGCTAGCGGCTAGCGGCTAGCGGCTAG 167
|||||
210 rAspSerGlyLysAspArgCysLeuHisGlnSerLysCysLysArgAlaLeuHisP 227
|||||
166 .. .. GGCACGGCGAGAGCTCGGG 148
|||||
227 rGlnVal .. AspProLysLeuLeuAspLysLeuHisCysLysThrPheHis 242
|||||
147 CGAGGGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTCGCTTC 98
|||||
243 GlnProAsnLysPhePheLysLeuValGlyArgThr 255
|||||
97 TCGCCACCGCTGATGGGTTCTCCATCATTCACCGCAAC 59

seq_name: /cn2_6/ptodata/2/ina/5a_311 x US-09-357-071-1
seq_documentation_block:
: Sequence 1, Application US/09457071
: Patent No. 6043091
: GENERAL INFORMATION:
: APPLICANT: Brett P. Monia
: APPLICANT: Lex M. Cowser
: TITLE OF INVENTION: ANTISENSE MODULATION OF LIVER GLYCOGEN PHOSPHORYLASE EXPRESSION
: FILE REFERENCE: RTS-0074
: CURRENT APPLICATION NUMBER: US/09/357,071
: CURRENT FILING DATE: 1999-07-19
: NUMBER OF SEQ ID NOS: 47
: SEQ ID NO 1
: LENGTH: 2828
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (114)..(2657)
US-09-357-071-1

alignment_scores:
Quality: 87.00 Length: 236
Ratio: 0.798 Gaps: 11
Percent Similarity: 45.186 Percent Identity: 21.610

alignment_block:
US-09-357-262-2_copy_53_311 x US-09-357-071-1 ..
Align seq 1/1 to US-09-357-071-1 from 1 to 2828

23 LeuGluMetLeuSerLeuHisProAspValAlaAlaAlaGluAsnGluVa 49
|||||
1119 AATCAAGTAAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1168
|||||
39 lHisPhePheAspTrpGluLleHisTyrSerGlnGlyLeuGlyTyrTrp 56
|||||
1169 GATTTTGTGTGATTTGAA .. 1187
|||||
56 eThrGlnMetProPheSerSerProHisGlnLeuThrValGluLysThr 72
|||||
1188 .. .. AATCAAGTAAATCAATCAATCAATCAATCAATCAATCAATCAAT 1229
|||||
73 ProAlaTyrPheThrSerProLysValProGluArqLleHisSerMetAs 89
|||||
1230 .. .. AATCAAGTAAATCAATCAATCAATCAATCAATCAATCAATCAAT 1277
|||||
89 nProThrIleArqLeuLeuLleLeuArqAspProSerGluArqVal 106
|||||

```



```

61 eSerSerProHisGlnLeuThrValGluLysThrProAlaTyrPheThrS 78
157 ACCAACAGAGTACTTGGCTTCAGCAAGTCAACCAATTCAC 198
78 eProLysValProGluArgIleHisSerMetAsnProThrIleArgLeu 94
199 NAGAAATTCAGATTCTATAATATGAAAGATATTTGTGATTC 246
95 LeuLeuIleLeuArgAspProSerGluArgValLeuSerAspTyrThrGI 111
247 CACAAATTTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 269
111 nValLeuTyrAsnHisLeuGlnLysHisLysProTyrProProTleuA 128
270 TATGGTTTGGGACAAC AATTGTCCTTTCCTTCCTTCCTTCCTTCCTTC 310
128 sLeuLeuMetArgAspGlyArgLeuAsnLeuAspTyrIleAlaLeuAsn 144
311 CTATCCACAAGATTTAAAGAGGTTTCTGATTTTCTGATTTTCTGATTTT 354
145 ArgSerLeuTyrHisAlaHisMetLeuAsnTrpLeuArgPhePheProLe 161
355 CAGAAGGCTGATCAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 387
161 uGlyHisIleHisIleValAspGlyAspArgLeuLeuLeuLeuLeuLeu 175
388 ATTTCAAGAGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 424
175 sProPheProGluIleGlnLysValGluArgPheLeuLysLeuSerPro 191
425 AATTCAGAGAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTA 474
192 GlnIleAsnAlaSerAsnPheTyrPheAsnLysThrLys.....Glyp 206
475 TATTTGGGGTGACACATTTGGTTCGTTCATCTTCCTCTCTCTCTCTCT 524
206 eTyrCys 208
525 CTATTC 531

```

seq_name: /cgn2_6/ptodata/2/ina/2A_comb.seq.us-08-441-139-17

```

seq_documentation_block:
; Sequence 17, Application US/08441119
; Patent No. 5773245
; GENERAL INFORMATION:
; APPLICANT: Wittrup, Dr. Karl D.
; APPLICANT: Robinson, Anne S.
; TITLE OF INVENTION: METHODS FOR INCREASING SECRETION OF
; TITLE OF INVENTION: RECOMBINANTLY EXPRESSED PROTEINS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,139
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/089,997
; FILING DATE: 06-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346

```

```

; REFERENCE/DOCKET NUMBER: 8646
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELETYPE: 230 901 SANS OR
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-441-139-17

alignment_scores:
Quality: 84.50 length: 166
Ratio: 1.006 Gaps: 8
Percent Similarity: 50.502 Percent identity: 25.301

alignment_block:
US-09-557-262-2_copy_53_311 x US-08-441-139-17
Align seq 1/1 to: US-08-441-139-17 from: 1 to: 2403
105 ValLeuSerAspTyrThrGln.....ValLeuTy 114
1186 GTTTCGCCCAATACGTCGAAAGCGGTTCGCTTTCGCTTTCGCTTTCGCTT 1245
114 rAsnHisLeuLysHisLysProTyrProProIleGluAspLeuLeuM 131
1236 CAATGAGAGAGAAATGGAAGATTAAGAGATTAAGAGATTAAGAGATTAAG 1285
131 eTyrAspTyrArg.....LeuAsnLeuAspTyrLysAla 142
1286 CCAAAAGACAGAGGCTTAATGAACCTTGTTCATGATGATGATGATGATGAT 1335
143 LeuAsnArgSerLeuTyrHisAlaHisMetLeuAsnTrpLeuArgPheP 159
1336 TTTCGCTGAGA CAGGCGGCAACCTGCAACATGAGGAGCAACAT 1376
159 eProLeuGlyHisIleHisIleValAspGlyAspArgLeuLeuLeuLeu 176
1377 CCTTTATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1420
176 rPheProGluIleGlnLysValGluArgPheLeuLysLeuSerPro 182
1421 GTTTCGCTGAGA CAGGCGGCAACCTGCAACATGAGGAGCAACAT 1467
193 Ile...AsnAlaSerAsnPheTyrPheAsnLysThrLysGlyPheTyrCy 208
1468 ATCGCTTTCGAGCTTCAGGCTATTCAGATTCGCTTTCGCTTTCGCTTTCG 1515
208 sLeuArgAspSerGlyLysAspArgCysLeuHisGluSerLysGlyArgA 225
1516 .. AAAGTGTATG 1525
225 rHisProGlnValAspProLysLeuLeuAspLysLeuHisGluTyrPhe 241
1526 CAGTCAAAATTCAG 1557
242 HisGlyProAsnLysLysPhePheLeuValIleArgThrPheAsp 257
1558 CAGAACTAAGATTCGCTTTCAGATTCGCTTTCAGATTCGCTTTCAGATTC 1605
seq_name: /cgn2_6/ptodata/2/ina/2B_comb.seq.us-09-056-556-183
seq_documentation_block:
; Sequence 183, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.

```



```

141 lysAlaIleuAsnArgSerLeuTyrHisAlaHisMetIleuAsnTrpIleuA 157
      :::::  |||||  :::::  |||||  |||||  |||||  |||||  |||||
1481 ATGGCATCTCTCGAGCTGAG      AGGTTGAGAGTGGATGAG 1441
      :::::  |||||  :::::  |||||  |||||  |||||  |||||  |||||
157 qPhePheProLeuGlyHisIleHisIleValAspGlyAspArgLeuIleA 174
      |  ::::  ::::  ::::  ::::  |||||  |||||  |||||  |||||
1440 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1391
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
174 rQAspProPhePro 178
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
1390 TCTCTCCATACCCC 1377

```

```

seq_name: /cgr2_6/ptodata/2/ina/5A_GOMB.seq US-08-727-034-2

```

```

seq_documentation_block:
; Sequence 2: Application US/08727034
; Patent No. 5665872
; GENERAL INFORMATION:
; APPLICANT: SATO, YASHUSHI
; APPLICANT: IWASAKI, AKIO
; APPLICANT: ARAI, KOICHI
; APPLICANT: YAMAZAKI, HIROYUKI
; TITLE OF INVENTION: NOVEL LDL RECEPTOR ANALOG PROTEIN AND
; TITLE OF INVENTION: THE GENE CODING THEREFOR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GELON, SPIVAK, McLELLAND, WATER & NEUSTADT.
; ADDRESSEE: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/727,034
; FILING DATE: 08-OCT-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 261440/1995
; FILING DATE: 09-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 102451/1996
; FILING DATE: 24-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: OHLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/BOOKET NUMBER: 80-0/9-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6961 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 178..6961
; OTHER INFORMATION: /note "Identification Method: S"
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 178..261
; OTHER INFORMATION: /note "Identification Method: S"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 262..6816

```

```

; OTHER INFORMATION: /function- "Nucleotides 262-6816
; OTHER INFORMATION: encode the mature peptide"
; OTHER INFORMATION: /note- "Identification Method: S"
US-08-727-034-2

alignment_scores:
      Quality: 79.50      Length: 172
      Ratio: 0.970      Gaps: 7
      Percent Similarity: 47.674      Percent Identity: 26.163

alignment_block:
US 09 557 262 2 copy 53 31l x US-08 727 034 2 rev
      Align seq 1/1 to reverse of: US-08-727-034-2 from: 1 to: 6961

      10 IleIleIleGlyValArgLysGlyThrArgAlaLeuLeuGluMetLe 26
          :::::  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      1992 CTCCTGTGTGGAGGCAAGATGGTGAAGACGCGTGTCTCTCTCTC 1951
          :::::  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      26 uSerLeuHisProAspValAlaAlaAlaGluAsnGluValHisPhePheA 43
          :::::  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      1950 GCGCGCTTCGCTGAGAGCGCGCATACACAAACAG 1917
          :::::  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      43 sPTP.....GluGluHisTyrSerGluGlyLeuGlyTyrTyrLeuThr 57
          |||  ::::  |||||  |||||  |||||  |||||  |||||  |||||
      1916 GATTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1867
          :::::  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      58 GluMetProPheSerSerProHisGluLeuThrValGluLysThrProAl 74
          ::::  ||||  ||||  ||||  ||||  ||||  ||||  ||||
      1866 GTACTTCAGTTCGCTGCTTCTCCATGCGCTGGCGCAAGGGGATGATGATG 1817
          :::::  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      74 dTTPheThrSerPheLysValProGluArgIleHisSerMetAsnProT 91
          :::::  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      1816 GCGCATGCTGCCCATGTATA.....G 1794
          :::::  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      91 hrIleArgLeuLeuLeuLeuArgAspProSerGluArgValIleuSer 107
          :::::  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      1793 TAGTGAGTCCAGGAGATGGCTCTGGGCAAGCTGTGCTGACACACTGTAGA 1744
          :::::  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      108 AspTyrThrGluValLeuTyrAsnHisLeuGluGlyHisLeuLysProL 154
          |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      1743 CATGTACACAGTTTGTCTTG.....GACGCCAAGTCTTTCCTCCACATG 1703
          :::::  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      124 rProIleGluAspLeuMetArgAspGlyArgLeuAsnLeuAspTyr 140
          |||||  ::::  |||||  |||||  |||||  |||||  |||||  |||||
      1762 AGAGAGTGGAGATATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1659
          :::::  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      141 LysAlaLeuAsnArgSerLeuTyrHisAlaHisMetIleuAsnTrpIleuA 157
          :::::  |||||  |||||  |||||  |||||  |||||  |||||  |||||
      1658 ATGGCATCTCTCGAGCTGAG      AGGTTGAGAGTGGATGAG 1618
          :::::  |||||  :::::  |||||  |||||  |||||  |||||  |||||
      157 qPhePheProLeuGlyHisIleHisIleValAspGlyAspArgLeuIleA 174
          |  ::::  ::::  ::::  ::::  |||||  |||||  |||||  |||||
      1617 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1568
          ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
      174 rQAspProPhePro 178
          ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
      1567 TCTCTCCATACCCC 1554

```

```

seq_name: /cgr2_6/ptodata/2/ina/5A_GOMB.seq US-08-727-034-2
seq_documentation_block:
; Sequence 1: Application US/08363300
; Patent No. 5780927
; GENERAL INFORMATION:
; APPLICANT: ZON, Leonard and Richardson, Paul
; TITLE OF INVENTION: Tbet Gene and Uses Thereof
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston

```



```

2949 TACAGCACCTCCATCCCCAGCCGCTGCTATTTCTCTTCTTCTATGGAAGAC 2900
134 GlyArgLeuAsn.LeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisA 150
2899 GTACAGCATATAATCACCAACCTTCCCAAGCATCAACAGGACAGGACGGGCA 2850
150 LeuHisMetLeuAsnTrpLeuArgPhe...PheProLeu..... 161
2849 GAGCGATGATCAGATGGATGAAGTTTTCATATCTGTGTTGGGCTGGACA 2800
162 .....GlyHisIleHisIleValAspIleArgLeuIleArg 174
2799 TAGAAGAACACAGGATCTCTCCAC..... 2776
174 AspProPheProGlu 179
2775 .GAGCCATTCCACAG 2761
```


234 euAspLysLeuHisGluTyrPheHisGlnProAsnLysIstySphSerPheLys 250
|||||
61 TCACATAAAATSCAGCAATATTTCATAGAGCAATAAACAAGTCTTTGGC 12

251 LeuValGly 253
|||||
11 CTGTGTGGC 3

seq_name: /SIDS5/qcudata/qeueseq/qeueseq-emb1/NA2001A.DAT.AAK16371

seq_documentation block:

ID AAK16371 standard; DNA; 924 bp.

✕✕

AC AAK16371;

XX

DT	05-NOV-2001	(first entry)	Human brain expressed single exon probe SEQ ID NO: 16362.
XX			
DE			
XX			
XX			Human; brain expressed exon, gene expression analysis, probe, microarray; Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer; ss.
KW			
KW			

XX

OS Homo sapiens

XX

PN WO200157275-A2.

XX

PD 09-AUG-2001.

XX

30-JAN-2001; 2001WO-US006667.

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

04-FEB-2000; 0000Z; 210810Z

PR 26-MAY-2000; 20:01US-0207456

PR 30-JUN-2000; 2000S-0608408.
DB 02-AUG-2000; 2000S-0625304.

PR 03-AUG-2000; 2000US-
PR 21-SEP-2000; 2000US-

PK 27-SEP-2002; 2000US-0216149.
 PR 04-OCT-2000; 2000GB-0014263.
 XX
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INT.
 XX penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI: 2001-483446/52.
 XX
 DR
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 PT

XX
XX

Example 4; SI

XX
XX

CC The present invention provides a number of single

CC probes which are derived from genomic sequences ex

CC brain. They

CC which may enable the diagnosis and improved treatment of

diseases such as Alzheimer's disease, multiple sclerosis,

epilepsy and cancers. The present sequence is one of the

761 GGCCTCTGGACACACTTGGCCCTACACATCATCAATGCGGCTGGCAAGGG 712

17 yGlyThrArgAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaA 34

711 CAGCAACAGAGGCACTGCTGGAGATGGTCTACATGACACGAGCGAGCG 662

34 IaAlaGluAsnGluValHisPhePheAspTrpGluGluHisItyrSerGln 50

661 CCGGACAAAGAGAGATGATTTTCAATGATGAGAGAGATATTAAGAGAG 612

51 GlyLeuGlyItyrItyrLeuGlnMetProPheSerSerProHisGlnLe 67

611 GGCCTTGGCGCGGACCTACAGCGCAATGGCTCTCTCTGGCCACACGAGCT 562

67 uThrValGluLysThrProAlaItyrPheThrSerProLysValProGluA 84

561 CACATCTGACAGAAATGCTGGGAGATATGACAGAGGACCAAGTGGCTGAG 512

84 rglIeHisSerMetAsnProThrIleArgLeuLeuLeuIleLeuArgAsp 100

511 GAGCTTATAGATATAGAGGCTGATGATGATGATGATGATGATGATGAT 462

101 ProSerGluArgValIleuSerAspItyrThrGlnValLeuItyrAsnHisLe 117

461 CCGCTGGGAGGCGGCGCTATCTGACATACACCAAGAGAGAGAGAGAGAT 412

117 uGlnLysHisLysProItyrProProIleGluAspLeuLeuMetArgAspG 134

411 GCAGAGAGACAAAGTGGTACCGGCTCCATCGAGAGAGTCTCTGGTGGGAG 362

134 IyArgLeuAsnLeuAspItyrLysAlaLeuAsnArgSerLeuItyrHisAla 150

361 GATGATTTAAATGCTGATATAGAGGCTGATAGAGGCTGATAGAGGCTG 312

151 HisMetLeuAsnTrpLeuArgPhePheProGluGlyHisIleHisIleVa 167

311 CACATCTACAAATGGCTGGCGTCTCTGGCGTGGCTGGCTGGCTGGCTGG 262

167 IAspGlyAspArgLeuIleArgAspProPheProGluIleGlnLysValG 184

261 GCATGGGCAAGGCTGATACAGAGAGAGGCTGCTGGCTGGCTGGCTGGCT 212

184 IuArgPheLeuLysLeuSerProGlnIleCasnAlaSerAsnPhenylItyrPhe 200

211 AGAGGTTCTAAAGTCTGGTGGTGAATATATGATGAGATCTTACATCTT 162

201 AsnLysItyrLysGlyPheTrpCysLeuArgAspSerGlyLysAsnArgG 217

161 AACCAAAATATGGCTTTATATGGCTGGGAGAGAGGCGGCGGCGGCGG 112

217 sLeuHisGluSerLysGlyArgAlaHisProGlnValAspProLysLeuL 234

111 CTATACATCAAGCTTAAAGGCGGAGGCTGATGCTGAGTGAAGTGGATGCCAAACTAC 62

234 euAspLysLeuHisClnItyrPheHisClnItyrAsnLysLysPheLys 250

61 TCATATATATATGATGATATTTTCATGAGGCAATATAGAGATTCCTCGAG 12

251 LeuValGly 253

11 CTGTGTGGC 3

[illegible]

seq_documentation_block:

11D AAK42122

XX

AC AK42122;

XX Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss
 XX Homo sapiens.
 US W0200157276 A2.
 PN 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-050066R
 XX 04-FEB-2000; 2000US-0180312.
 XX 26-MAY-2000; 2000US-0207456.
 PP 30-JUN-2000; 2000US-0608408.
 PK 03-AUG-2000; 2000US-0632466.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Bank D9;
 XX WPI; 2001 48R400/53
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX Example 4; SEQ ID NO: 16679; 658pp + Sequence listing; English.
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 XX SQ Sequence 924 BP: 169 A: 252 C: 310 G: 104 T: 0 other:

alignment_scores:

Quality: 1277.00 Length: 253
 Ratio: 5.129 Gaps: 0
 Percent Similarity: 98.419 Percent Identity: 92.095
 alignment_block:
 US-09-557-262-2_copy_53_311 x AAK4122/rev ..
 Align seq 1/1 to reverse of: AAK42122 from: 1 to: 924
 1 GlySerThrGlnGlnLeuProGlnThrIleIleIleGlyValArgLysG1 17
 761 GGCCTGCCCCAGGAGTTCGGGACACCAATCAATCGGGTGGCGAAGGG 712
 17 yGlyThrArgAlaLeuLeuGluMetLeuSerLeuHisProAspValAla 34
 711 GGGTAAAGAGAGTAAATGAGAGTAAATGAGAGTAAATGAGAGTAA 662
 34 laAlaGluAsnGluValHisPhePheAspTrpGluGluHisTyrSerGln 50
 661 CCGGCGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 512
 51 GlyLeuGlyTrpTyrLeuThrGlnMetProPheSerSerProHisGlnLe 67
 611 GGGTGGGGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 562
 67 nThrValGluGlyThrProAlaGlyThrSerSerSerSerSerSerSer 84
 561 CAAGTGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 512
 84 rglHisSerMetAsnProThrIleArgGluLeuLeuLeuArgAsp 100

511 GAGTCTACGCAATGAACCGCTCCATCCGGCTCTCTCTCTCTCTCTCT 462
 101 ProSerGluArgValLeuSerAspTyrThrGluValLeuTyrAsnIleLe 117
 461 CCGTGGAG 412
 117 uGlnHisLysProTyrProGlnIleGluAspLeuLeuMetArgAspG 134
 411 GGAAGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362
 134 lyArgLeuAspLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisAla 150
 361 GCAAGCTCAATGTGGACTACAAAGGCGGAGAGAGAGAGAGAGAGAG 412
 151 HisMetLeuAsnTrpLeuArgPhePheProLeuGlyHisIleIleVal 167
 311 CAATAGAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 262
 167 AspGlyAspArgLeuIleArgAspProPheProGlnIleGlnLysValG 184
 261 GGAAGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 212
 184 LuArgPheLeuLysLeuSerProGlnIleAsnAlaSerAspPheTyrPhe 200
 211 AGAGCTTCTTAAAGGCTGCGGCGGAGAGAGAGAGAGAGAGAGAGAG 162
 201 AsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyLysAspArgG 217
 161 AAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 112
 217 LeuHisGluSerLysGlyArgAlaHisProGlnValAspProLysLeuL 234
 111 GTTAAAG 62
 234 euAspLysLeuHisGlyTyrPheHisGluProAsnLysLysPhePheLys 250
 61 TCATTAACATCCAGCAATATTTTCATGAGCCAAATTAAGAAAGTCTCT 12
 251 LeuValGly 253
 11 CTTGTTGGC 3
 seq_documentation_block:
 ID AAI48186 standard; DNA: 924 BP.
 XX
 AC AAI48186;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #16872 used to measure gene expression in human placenta sample.
 XX
 KW Probe: microarray; human, placenta, antenatal diagnosis;
 KW genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-050066R
 XX
 PP 04-FEB-2000; 2000US-0180312.
 PP 26-MAY-2000; 2000US-0207456.
 PP 30-JUN-2000; 2000US-0608408.
 PP 03-AUG-2000; 2000US-0632466.
 PP 21-SEP-2000; 2000US-0234687.
 PP 27-SEP-2000; 2000US-0236359.
 PP 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE) MOLECULAR DYNAMICS INC.

Claim 5; Page 152-153; 158pp; English.

PS The present sequence is human drug metabolising enzyme (DME)-5 cDNA.
 XX DME polypeptide, polynucleotide and modulators are useful for
 CC diagnosis, treatment and prevention of autoimmune/inflammatory,
 CC cell proliferative, developmental, endocrine, eye, metabolic,
 CC and gastrointestinal disorders, including liver disorders.
 CC The autoimmune/inflammatory disorders treatable include
 CC AIDS, adult respiratory distress syndrome, Addison's disease,
 CC allergic anaphylaxis, asthma, atherosclerosis, osteoporosis, autoimmune
 CC hemolytic anaemia, autoimmune thyroiditis, Crohn's disease, atopic
 CC dermatitis, diabetic mellitus, Graves' disease, glomerulonephritis,
 CC rheumatoid arthritis, scleroderma, systemic lupus erythematosus,
 CC systemic sclerosis, ulcerative colitis, haemodialysis and uveitis,
 CC viral, bacterial, fungal, parasitic, protozoan, helminthic infections
 CC and trauma, and cell proliferative disorders such as cancer, actinic
 CC keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis,
 CC hepatitis and psoriasis. Developmental disorders include anaemia, renal
 CC tubular acidosis, epilepsy, hypothyroidism and cataract, and endocrine
 CC disorders include disorders of hypothalamus and pituitary, disorders
 CC associated with hypopituitarism, including sarcoidosis, diabetes
 CC insipidus, hypogonadism, disorders associated with hypothyroidism
 CC including goitre, acute thyroiditis, Graves' disease, disorders
 CC associated with hyperparathyroidism, pancreatic disorders such as type I
 CC or type II diabetes mellitus, disorders associated with adrenals such as
 CC hyperplasia, Cushing's disease, endometriosis, infertility.
 CC hypergonadal disorders, and gynaecomastia. Eye disorders include
 CC conjunctivitis, keratitis, glaucoma and macular degeneration, and
 CC metabolic disorders include diabetes, cystic fibrosis, goitre,
 CC hypercholesterolaemia, hypothyroidism, hyperlipidaemia, lysosomal storage
 CC diseases, obesity, phenylketonuria and hypocalcaemia. Also the molecules
 CC are useful for treating gastrointestinal disorders such as dysphagia,
 CC gastritis, peptic ulcer, cholelithiasis, cirrhosis, hepatitis,
 CC hyperbilirubinaemia, constipation, diarrhoea, jaundice, Wilson's disease,
 CC thrombosis and hepatic tumours. The DME polypeptide is also useful for
 CC screening its agonist or antagonist.

50 Sequence 2115 bp, 651 A, 408 C, 438 G, 618 T, 0 other;

alignment_scores:
 Quality: 755.96 Length: 258
 Ratio: 3.564 Gaps: 2
 Percent Similarity: 82.171 Percent Identity: 53.488

alignment_block:
 US-09-557-262-2_copy_53_311 x AAD24010 ..

Align seq 1/1 to: AAD24010 from: 1 to: 2115

4 GlnGlnLeuProGlnThrIleIleIleGlyValArgLysGlyGlyThrAr 20
 438 CAGCAGCCGCCAAGGCCATATATCATTTGGCGTCAGCAAGAGCCACAAC 487
 20 aAlaLeuLeuGluMetLeuSerLeuLeuHisProAspValAlaAlaGluA 37
 488 GATCTCTGTTCGAAATGCTGAGACCTACATCGCGGAGTAGTCAAAAGCCCTCTC 537
 17 snGluValHisPheAspTrpGlnGluHisIstyrSerGlnGlyLeuGly 53
 538 AAGAAATCCACTTTTTTGATATGATGAGAAATTAIGTAAGGGGCAATTGAG 587
 54 TrpIstyrLeuThrGlnMetProPheSerSerProHisGlnLeuThrValGI 70
 588 TGGATATAGCAAAAGATGCTTTTTTCTACCTCACTCAACATCAATCAATCA 637
 70 uLysThrProAlaIstyrPheThrSerProLysValProGluAlaIleHis 87
 638 AAACAGCCCACTATATTTTATATATATATATATATATATATATATATATAT 687
 87 crMetAsnProThrIleArgLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 103
 688 AATATGAAAT 737

104 ArgValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLysHis 120
 738 AGAGCTATTTCTGATATATCTAGGTCAGGCGCAAGGAGGAGGAGGAGGAG 787
 120 stYsPTrpTrpProProIleGluAspLeuLeuMetArgAspGly.....A 145
 788 CAAATCTATTACAAAGTTTGGAGAGCTGGGCAATAGACCAATTAATACAGGG 847
 135 rGLeuAsnLeuAspTyrIstyrIstyrAlaLeuAsnArgSerLeuIstyrHisAlaHis 151
 838 AAGTGAAACAAAAAIAAAGAGAGTAAGAACAGCAATCACTACAAACAAAT 887
 152 MetLeuAsnTrpLeuAlaPhePheProLeuGlyIstyrIstyrIstyrIstyr 168
 888 CTGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 947
 168 pGlyAspArgLeuIleArgAspProPheProGluIleGlnLysValGluA 185
 938 TGAAT 987
 185 rPheLeuLysLeuSerProGlnIleAsnAlaSerAsnPheIstyrPheAsn 201
 988 AGTTCTCAATATCTGCTCCCAAGGATAAGTCAATACAAATTTATATTTCAAT 1047
 202 LysThrLysLysPheTyrCysLeuArg.....AspSerGlyLysAspArgGly 217
 1038 CCAACACAGAGGCTTTTACCTTCGCGCTTTAAATATATATATATATATATAT 1087
 217 sLeuHisGlnSerLysGlyArgAlaHisProGluValAspProLysLeuL 244
 1088 GTTGAG 1147
 234 euAspLysLeuHisSerLysIstyrPheHisSerLeuProAsnLysLysPhePheLys 250
 1138 TTACTAATTCGGCAATTCCTTCATCTCTTTTAAATCAAAAAATTTTACGAG 1187
 251 LeuValGlyAlaThrPheAspTrp 258
 1188 ATCACTGGGAGGACATTCGAACCTGG 1211

seq_name: /SILS5/seqdata/geneseq/geneseq_embd/NA2001A.DAT:ABA44908
 seq_documentation_block:
 ID ABA44908 standard; DNA; 452 bp.
 XX
 AC ABA44908;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Human breast cell single exon nucleic acid probe #3604.
 XX
 KW Human, microarray, single exon probe, gene expression; breast;
 KW disease; cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200157271-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001W9-US00662.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0634566.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000CB-0324263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI: 2001-496933/54.
 XX New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -
 XX
 PS Claim 1: SEQ ID NO 3603; 427pp + sequence listing; English.
 XX The invention relates to a spatially addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and H1 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or producing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 452 BP; 86 A; 114 C; 155 G; 97 T; 0 other;

alignment_scores:
 Quality: 748.00 Length: 150
 Ratio: 5.123 Gaps: 0
 Percent Similarity: 97.333 Percent Identity: 89.333

alignment_block:

US-09-557-262-2_COPY_54_411 x ABA44406/rev

Align seg 1/1 to reverse of: ABA444908 from: 1 to: 452

37 AsnGluValHisPheAspTrpGluGluHisTyrSerGlnGlyLeuGI 53
 |||||
 451 ACGAGGTCACCTTTCTGACTGGGAGGAGATTACAGCCAGCGCTGGG 402
 53 YTrpTyrLeuThrGlnMetProPheSerSerProHisLeuThrValG 70
 |||||
 401 CTGCTACCTTCACACATCCCTCTCTCTGCTGCTGCTGCTGCTGCT 352
 70 LuLysThrProAlaTyrPheThrSerProTyrValProGluArqIleHis 86
 |||||
 351 AGAAGAGCGCGCGCGCTATTATATGTTGGGAAATGCTGCTGCTGCTG 402
 87 SerMetAsnProThrIleArgLeuLeuLeuIleLeuArgAspProSerGI 103
 |||||
 301 AGATGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 252
 103 uArqValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLysH 120
 |||||
 251 CGCGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 202
 120 isLysProTyrProProIleGluAspLeuLeuMetArqAspGlyArgLeu 136
 |||||
 201 ACAAGGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 152
 137 AsnLeuAspTyrTyrAlaIleAsnArgSerLeuTyrHisAlaHisMetLe 153
 |||||
 151 AATGLeuValHisPheAspTrpGluGluHisTyrSerGlnGlyLeuGI 102
 153 uAsnTrpLeuArqPheProLeuGlyHisIleIleHisIleValAspGlyA 176
 |||||
 101 GAACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 92

170 SPArgLeuIleArgAspProPheProGluIleGlnLysValGluArgPhe 186
 |||||
 51 ACGAGGTCACCTTTCTGACTGGGAGGAGATTACAGCCAGCGCTGGG 2

seq_name: /SI055/qcdata/geneseq/geneseq_emb1/NA2001A.DAT.ABA55377

seq_documentation_block:

ID ABA55377 standard; DNA; 452 BP.

AC ABA55377;

DT 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #3682.

KW Human; foetal liver; gene expression: single exon nucleic acid probe; ss.

OS Homo sapiens.

PN W0200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001W01500669.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0612366.

PR 21-SEP-2000; 2000US-0134887.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000EP-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI: 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human fetal liver -

PS Claim 1: SEQ ID NO 3682; 639pp + sequence listing; English.

CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention

CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 452 BP; 86 A; 114 C; 155 G; 97 T; 0 other;

alignment_scores:

Quality: 748.00 Length: 150

Ratio: 5.123 Gaps: 0

Percent Similarity: 97.333 Percent Identity: 89.333

alignment_block:

US-09-557-262-2_COPY_53_311 x ABA55377/rev

Align seg 1/1 to reverse of: ABA55377 from: 1 to: 452

37 AsnGluValHisPheAspTrpGluGluHisTyrSerGlnGlyLeuGI 53

|||||

451 ACGAGGTCACCTTTCTGACTGGGAGGAGATTACAGCCAGCGCTGGG 402

53 YTrpTyrLeuThrGlnMetProPheSerSerProHisLeuThrValG 70

```

|||||
401 CAGTATCTCAAGAAATAGATTTCTTGGAAATATTAATATG 352
70 IulysThrProAlaTyrPheSerProGlyValProGluArgHis 86
|||||
351 AAGAGAGCGCCGCTATTATATGAGGAAAGAGGCTGAAATAT 402
87 SerMetAsnProThrIleArgLeuLeuLeuIleLeuArgAspProSerG1 103
|||||
301 AGCAATCAACCCCTCCATCCGCTGCTGCTCAATCCCTCCGAGACCGCTCGGA 252
103 uArgValLeuSerAspTyrThrGluValLeuTyrAsnHisLeuGluLysH 120
|||||
251 GAGATATATATATATATATATATATATATATATATATATATATATATAT 202
120 IulysProTyrProIleGluAspLeuLeuMetArgAspGlyArgLeu 136
|||||
201 AATAGAGCTTACCTGATATATATATATATATATATATATATATATATAT 152
137 AsnLeuAspTyrLysAlaLeuAsuArgSerLeuTyrHisAlaHisMetLe 153
|||||
151 AATGCGCACTAATAGAGCCCTTACCGTAGCTGATACAGAGTACATGCA 102
153 uAsnTrpLeuArgPheProLeuGlyHisIleHisIleValAspGlyA 170
|||||
101 GAACATGGCTGAGATTTTTCAGGCTGGAGCAATGCAATTTGGACGGCG 52
170 sArgLeuIleArgAspProPheProGluIleGlnTysValGluArgPhe 186
|||||
51 AGCGGCTCATAGGAGCCCTTTCCTGAGATCCAAAAGGTCGAGAGGTTTC 2

```

seq_name /SUS5/seqdata/geneseq/geneseq-rmb1/NA25103 [AT-ARA25103]

seq_documentation_block:

ID_ABA25103 standard; DNA, 452 BP.

AC ABA25103;

XX 23-JAN-2002 (first entry)

DE Probe #3569 for gene expression analysis in human heart cell sample

KW Human; gene expression, heart, microarray, vascular system; probe;

KW cardiovascular disease; hypertension; cardiac arrhythmia;

KW congenital heart disease, SS

XX Homo sapiens.

XX W0200157274 A2.

XX 09-AUG-2001.

XX 30-JAN-2001 - 2001W0 080666.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0256459.

XX 04-OCT-2000; 2000CB 0624264.

XX (MOLEC.) MOLECULAR DYNAMICS INC.

XX Penn SC, Hanzel DK, Chen W, Rank DK,

XX WPI; 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX hearts -

XX Claim 1; SEQ ID No 3569; 530pp; English.

XX

CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe; the probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging,
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from Wipo
 CC at http://wipo.int/pub/published_pat_sequences.

XX
 XX SeqAccess 452 BP, 85 A, 114 C, 155 G, 97 T, 0 other;

alignment_scores:

quality: 748.00 Length: 150

Ratio: 5.123 Gaps: 0

Percent Similarity: 97.333 Percent Identity: 89.333

alignment_block:

US-09-557-262-2_COPY_53_311 x ABA25103/rev

Align seq 1/1 to reverse of: ABA25103 from: 1 to: 452

37 AsnGluValHisPheAspTrpGluGluHisTyrSerGlnGlyLeuG1 53

|||||

451 AACAGAGCGCCGCTATTATATGAGGAAAGAGGCTGAAATAT 402

53 YTrpTyrLeuThrGlnMetProPheSerSerProHisGlnLeuThrValG 70

|||||

401 CAGTATCTCAAGAAATAGATTTCTTGGAAATATTAATATG 352

70 IulysThrProAlaTyrPheSerProGlyValProGluArgHis 86

|||||

351 AGAGAGCCCTTACCTGATATATATATATATATATATATATATATATATAT 402

87 SerMetAsnProThrIleArgLeuLeuLeuIleLeuArgAspProSerG1 103

|||||

301 AGATAGAGCCCTTACCTGATATATATATATATATATATATATATATATATAT 252

103 uArgValLeuSerAspTyrThrGluValLeuTyrAsnHisLeuGluLysH 120

|||||

251 GAGAT 202

120 IulysProTyrProIleGluAspLeuLeuMetArgAspGlyArgLeu 136

|||||

201 AATAGAGCTTACCTGATATATATATATATATATATATATATATATATATAT 152

137 AsnLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisAlaHisMetLe 153

|||||

151 AATGCGCACTAATAGAGCCCTTACCGTAGCTGATACAGAGTACATGCA 102

153 uAsnTrpLeuArgPheProLeuGlyHisIleHisIleValAspGlyA 170

|||||

101 GAACATGGCTGAGATTTTTCAGGCTGGAGCAATGCAATTTGGACGGCG 52

170 sArgLeuIleArgAspProPheProGluIleGlnTysValGluArgPhe 186

|||||

51 AGCGGCTCATAGGAGCCCTTTCCTGAGATCCAAAAGGTCGAGAGGTTTC 2

seq_name /SUS5/seqdata/geneseq/geneseq emb1/NA2001A.001:AAK03618

seq_documentation_block:

ID_AAK03618 standard; DNA, 452 BP.

XX

XX AAK03618;

XX 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe SEQ ID No: 3609.

XX

KW Human, brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer; ss.
 OS Homo sapiens.
 XX
 XX
 PW WO200157275-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00667
 XX
 PR 04-FEB-2000; 2000US-0180312
 PR 26-MAY-2000; 2000US-0207456.
 PP 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PP 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US 0236359.
 PR 04-OCT-2000; 2000GB 0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SQ, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488900/53.
 XX
 PT Single exon nucleic acid probes for analyzing gene expression in human
 PT brains.
 XX
 PS Example 4; SEQ ID NO: 3609; 650pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is one of the probes of the
 CC invention.
 XX
 SQ Sequence 452 nt; 86 A, 134 C, 155 G, 97 T, 0 other.

alignment_scores:
 Quality: 748.00 Length: 150
 Ratio: 5.123 Gaps: 0
 Percent Similarity: 97.333 Percent Identity: 89.333

alignment_block:
 US-09-557-262-2_copy_53_311 x AAK03618/rev ..
 Align seq 1/1 to reverse of: AAK03618 from 1 to 452

37 AsnGluValHisPheAspTrpGluTyrSerGlyLeuG1 53
 451 AACGAGTCCACCTTCCTGACCTGCGACGACGATACAGCCACGCTTGGG 492
 53 YrrpYrLeuThrGluMetProPheSerSerProHisGluThrValG 70
 401 CTGATGATACCTGATGATGATGATGATGATGATGATGATGATGATG 42
 70 LuysThrProAlaTyrThrSerProTyrValProGluArgGlyHis 86
 351 AAGAGAGCGCGCGGATTCATGATGATGATGATGATGATGATGATGATG 402
 87 SerMetAsnProThrIleArgLeuLeuLeuLeuLeuLeuLeuLeu 103
 301 AGCATGAGACCTTCCTGACCTGCGACGACGATACAGCCACGCTTGGG 252
 103 uArgValLeuSerAspTyrThrGluValLeuTyrAsnHisLeuGlu 120
 251 GAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 202
 120 lslyProTyrProThrIleGluAspGluLeuMetArgAspTyrArg 136

101 AATGTCATCAAGACGCTTCAAGCTGACCTTACGACGACGACGACGAC 102
 153 uAsnTrpLeuArgPheProLeuGlyHisIleHisIleValAspGlyA 170
 101 GAACAGTGGTGTCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 52
 170 sPArgLeuIleArgAspProPheProGluTyrGluTyrValGluArgPhe 186
 51 ACGGCTGATCAAGGACGCTTCAAGCTGACCTTACGACGACGACGACG 2
 seq_name: /SIDES/qc4data/geneSeq/geneSeq.cab/NA5001A.LAI.AAK29079

seq_documentation_block:
 ID AAK29079 standard, DNA, 452 BP.
 XX
 AC AAK29079;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human bone marrow expressed single exon probe SEQ ID NO: 3636.
 XX
 KW Human, bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200157275-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00668.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US 0234687.
 PR 27-SEP-2000; 2000US 0236359.
 PR 04-OCT-2000; 2000GB 0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SQ, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488900/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow.
 XX
 PS Example 4; SEQ ID NO: 3636; 658pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 XX
 SQ Sequence 452 BP; 86 A; 114 C; 155 G; 97 T; 0 other;

alignment_scores:
 Quality: 748.00 Length: 150
 Ratio: 5.123 Gaps: 0
 Percent Similarity: 97.333 Percent Identity: 89.333

alignment_block:
 US-09-557-262-2_copy_53_311 x AAK29079/rev ..

Align seq 1/1 to reverse of: AAK29079 from: 1 to: 452

```
37 AsnGlnValHisPhePheAspTrpGlnGlnHisIleTyrSerGlnGlyLeuG1 53
|||||
451 AAGGAGGTGCAATTCATGAGTGGAGAGATTAAGAGAGATTTGAG 402
|||||
53 YTrpTyrLeuThrGlnMetProPheSerSerProHisIleLeuThrValG 70
|||||
401 CTGGTACCTTCACCCACATGCTCTTCTCTCTCTCTCTCTCTCTCTCTCT 352
|||||
70 LeuLysThrProAlaTyrPheThrSerProLysValProGluArqIleHis 86
|||||
351 ACAAGACCTCCGAGGATTCATGAGTGGAGAGATTAAGAGAGATTTGAG 302
|||||
87 SerMetAsnProThrIleArgLeuLeuLeuLeuLeuLeuLeuLeuLeu 103
|||||
301 AGCATGAACTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 252
|||||
103 uArqValLeuSerAspTyrThrGlnValLeuTyrAsnHisIleLeuGlnLysH 120
|||||
251 GCGCGTCTATCTCATACACCCCAAGCTCTCTCTCTCTCTCTCTCTCTCT 202
|||||
120 IsLysProTyrProProIleGluAspLeuLeuMetArqAspGlyArqLeu 136
|||||
201 ACAAGCTTAACTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 152
|||||
137 AsnLeuAspIleLysAlaLeuAsnArgSerLeuTyrHisAlaHisMetLe 153
|||||
151 AATGTGAGTATTAAGAGTATTAAGAGTATTAAGAGTATTAAGAGTATTA 102
|||||
153 uAsnTrpLeuArqPheProLeuGlyHisIleHisIleValAspGlyA 170
|||||
101 CAACCTGGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 52
|||||
170 spArqLeuIleArqAspProPheProGluIleGlnLysValGluArqPhe 186
|||||
51 AGCGCTGATTAAGAGTATTAAGAGTATTAAGAGTATTAAGAGTATTAAG 2
```

OM of US 09-557-262-4_copy_49_307 Loc. GenBank.* out_format: .pfs
 Date: Jul 17, 2002 5:56 AM
 About: Results were produced by the GenCore software, version 4.5.
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE=frame.p2n.model -DEV=xlp
 -U/Cp2_1/Uspt0.spool/US09557262-4_copy_49_307 -runal 1-073002-161204-25577/arg_query fastal_1.1568
 -DB=GenEmbl -OFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -DOPEXT=0.000 -DOPEXT=0.000
 -GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=0.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cgi -LIST=45 -DOALIGN=200 -TUR_SCORE=pcr
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTENT.pfs
 -NORM=ext -HEAPSIZ=500 -MINFN=0 -MAXFN=2000000000
 -USPR=US09557262-4_copy_1-11342 NCPU=5 -CPU=3 -LONGLOC
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-557-262-4_copy_49_307

Query length: 259

Database: GenBank.*

Database sequences: 1797656

Database length: 187333701

Search time (sec): 7102.980000

score_list:

Sequence	Strd	Orig	Zscore	Escore	Len	Documentation
gb_pr:AF019386	+	1415.00	2604.31	1.1e-136	1305	AF019386 Homo sapiens heparan
gb_pr:AF073991	-	1415.00	2556.39	5.0e-134	156838	AF073991 Homo sapiens BA' cell
gb_pr:AC005230	+	1198.00	2542.37	2.8e-133	163298	AC005230 Homo sapiens chromo
gb_to:AC009133	+	1418.00	2422.02	1.5e-126	1657	HC009133 Mus musculus, heparan
gb_to:AF019385	+	1418.00	2421.85	1.5e-126	1685	AF019385 Mus musculus, heparan
gb_hlg:AC084416_3	+	1318.00	2480.20	3.2e-124	110800	Contiguation (4 of 6) of AC0
gb_hlg:AC098713	+	1318.00	2473.29	7.9e-124	220128	AC098713 Mus musculus chromo
gb_to:AF177430	+	1309.00	2411.02	6.2e-126	936	AF177430 Rattus norvegicus 3-O-
gb_pat:AX317978	+	772.50	1407.81	4.7e-70	2115	AX317978 Sequence 18 from patc
gb_to:AL355498	+	772.50	1364.35	1.2e-67	165549	AL355498 Human DNA sequence
gb_om:CA389534	+	667.00	1226.14	6.2e-60	519	AL389534 Canis familiaris mRNA
gb_pr:AF105374	+	598.00	1084.87	4.6e-52	1468	AF105374 Homo sapiens heparan
gb_to:AF105375	+	598.00	1084.87	4.6e-52	1968	AF105375 Homo sapiens heparan
gb_to:AF168992	+	585.00	1064.28	6.4e-51	1666	AF168992 Mus musculus D-glycos
gb_pr:AF105377	+	585.00	1060.44	1.1e-50	2032	AF105377 Homo sapiens heparan
gb_pr:AF105376	+	585.00	1058.40	1.4e-50	2546	AF105376 Homo sapiens heparan
gb_pr:AF023723	+	585.00	1057.22	1.6e-50	2808	AF023723 Homo sapiens cDNA F13
gb_pr:AF105378	+	555.00	1003.96	1.5e-47	2211	AF105378 Homo sapiens heparan
gb_hlg:AC017305	+	460.50	790.72	1.1e-45	94745	AC017305 Drosophila melanogas
gb_to:AC008337	+	456.50	786.42	2.2e-45	100000	AC008337 Drosophila melanog
gb_to:AF003511	+	460.50	778.39	5.4e-35	343807	AF003511 Drosophila melanog
gb_hlg:AC027132	+	147.00	760.44	5.4e-34	186752	AC027132 Homo sapiens chromo
gb_pr:AF003511	+	147.00	759.44	6.3e-34	188636	AF003511 Homo sapiens chromo
gb_hlg:AC003359	+	440.00	744.88	4.0e-33	219502	AC003359 Mus musculus clone
gb_hlg:AC088715	+	440.00	743.57	4.7e-33	219269	AC088715 Mus musculus chromo
gb_pr:AC094511	+	438.00	744.46	4.2e-33	189234	AC094511 Homo sapiens chromo
gb_hlg:AC021445	+	428.50	724.25	5.6e-32	203621	AC021445 Mus musculus clone
gb_to:CHL45803	+	426.50	748.66	2.5e-33	13172	AF098940 Caenorhabditis eleg
gb_to:AF098940	+	426.50	741.53	6.1e-33	24810	AF098940 Caenorhabditis eleg
gb_pr:AF006640	+	425.00	726.88	4.0e-32	81579	AF006640 Homo sapiens 16p13.3
gb_pr:HS430A6	+	423.00	720.07	2.7e-32	41255	AL331723 Human DNA sequence 1
gb_hlg:AL603889	+	421.00	735.84	5.9e-31	320096	AL603889 Mus musculus chromo
gb_pr:AC005363	+	420.00	718.46	1.2e-31	75108	AC005363 Homo sapiens chromo
gb_hlg:AC012180	+	420.00	706.93	5.4e-31	251997	AC012180 Homo sapiens chromo
gb_pr:AC005375	+	419.00	709.74	3.6e-31	149339	AC005375 Homo sapiens chromo
gb_pr:AC005374	+	419.00	708.61	4.2e-31	166687	AC005374 Homo sapiens chromo
gb_to:CELF32B10	+	365.50	624.13	2.1e-26	38024	U04990 Caenorhabditis elegans
gb_pr:AB036429	+	351.50	622.45	2.7e-26	3359	AF046429 Homo sapiens NUS1A mR
gb_to:AF175689	+	347.00	610.14	1.4e-25	4947	AF175689 Drosophila melanogast
gb_to:AB036838	+	346.50	615.13	6.7e-26	2732	AB036838 Mus musculus NUS1A mR

gb_pr:AF076005 1 342.50 506.85 1.3e-25 2478 1 AF076005 Homo sapiens hepar
 gb_pr:AF074924 1 342.50 506.18 5.1e-26 3198 1 AF074924 Homo sapiens hepar
 gb_to:AC004657 1 338.50 567.07 4.7e-24 76618 1 AC004657 Drosophila melano
 gb_hlg:AC018326 1 338.50 563.42 5.1e-24 110483 1 AC018326 Drosophila melano
 gb_to:AF009028 1 334.50 557.89 1.8e-22 162168 1 AF009028 Drosophila melan
 seq_name: gb_pr:AF019386
 seq_documentation_block:
 locus AF019386 1305 bp mRNA linear F01 14-NOV-1997
 DEFINITION Homo sapiens heparan sulfate 3-O-sulfotransferase-1 precursor
 ACCESSION AF019386
 VERSION AF019386.1 GI:2618972
 KEYWORDS
 SOURCE human
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1305)
 AUTHORS Shworak N.W., Liu J., Fritze L.M.S., Schwartz J.J., Zhang L.,
 Loefer D. and Rosenberg P.D.
 TITLE Molecular cloning and expression of mouse and human cDNAs encoding
 heparan sulfate D-glucosaminyl 3-O-sulfotransferase
 JOURNAL Biol. Chem. 272 (44), 28009-28019 (1997)
 MEDLINE 98010647
 REFERENCE 2 (bases 1 to 1305)
 AUTHORS Shworak N.W., Liu J., Fritze L.M.S., Schwartz J.J., Zhang L.,
 Loefer D. and Rosenberg P.D.
 TITLE Direct Submission
 JOURNAL Submitted (14-AUG-1997) Biology, MIT, 31 Ames Street, Cambridge, MA
 02139, USA
 FEATURES
 source 1..1305
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="Isolated from lambda Triplex Brain cDNA library
 from Clontech"
 gene 1..1305
 /gene="3OST1"
 5'UTR 1..118
 /gene="3OST1"
 sig_peptide 119..178
 /gene="3OST1"
 CDS 119..1042
 /gene="3OST1"
 /function="rate limiting enzyme for synthesis of
 anticoagulant heparan"
 /note="heparan sulfate sulfotransferase; 3-OST-1;
 interluminal Golgi resident protein (retention mechanism
 unknown)"
 /codon_start=1
 /product="heparan sulfate 3-O-sulfotransferase-1
 precursor"
 protein_id="AA084488.1"
 /db_xref="GI:2618973"
 translation="MAALGLGAVLVAGLVFSAELGQVELLRKAGTLDQDVROG
 VAPNSACGLPTIIIGVKGFTFALLEMLSLPEVAAAEVHFIFWEEHYSLGAW
 YLSQMFSPHOLITVEKTPAYSPKPFVYVSMNPISLLILIDPSERVLSYTOV
 FYNMGKKRPVSLPEHIVGGLAVAKNPISLYHVHMNMILFFPLRHIVDGD
 ELKPLFELKGVFSEFKLSQSNASMPFNKIKGIVYLPISQKPLHFSKQAHVQ
 VVPLKLKLHFHFHPKPKFPELVGRTHHW"
 179..1039
 /gene="3OST1"
 /product="heparan sulfate 3-O-sulfotransferase-1"
 263..268
 /gene="3OST1"
 /note="encodes potential N-linked site"
 279..1039
 /gene="3OST1"
 /note="encodes presumptive sulfotransferase catalytic
 domain"
 527..535
 misc_signal

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/qsc>

SOURCE INFORMATION:

The RPO1-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Moon, P. Y., Zhao, F., Frengsen, E., Taton, M., Catasuse, J. J., and de Jong, P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletier de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: PAGES.0 NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-669F3; the clone sequenced to the right is AC006230. Actual start of this clone is at base position 1 of RP11-512I20; actual end is at base position 159838 of RP11-512I20.

FEATURES

URFS
source

Location/Qualifiers

```
1. .159838
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
```

repeat_region	2. .775	/rpt_family="RPCI-11"
repeat_region	266. .317	/rpt_family="MER1_type"
repeat_region	770. .792	/rpt_family="T-rich"
repeat_region	845. .2469	/rpt_family="AT-rich"
repeat_region	2470 .2766	/rpt_family="L1"
repeat_region	2769 .3291	/rpt_family="ALu"
repeat_region	3321 .3441	/rpt_family="MER1_type"
repeat_region	3628 .3655	/rpt_family="L1"
repeat_region	4297. .4409	/rpt_family="AT-rich"
repeat_region	4815. .5120	/rpt_family="MIR"
repeat_region	5094 .5114	/rpt_family="ALu"
repeat_region	5203. .5546	/rpt_family="(A)n"
repeat_region	8632. .8739	/rpt_family="L2"
repeat_region	9549. .9890	/rpt_family="GA-rich"
repeat_region	10107. .10349	/rpt_family="ERV1"
repeat_region		/rpt_family="MIR"

repeat_region	11853..12152	/rpt_family="MaLR"
repeat_region	12289..12271	/rpt_family="L1"
repeat_region	12305..12390	/rpt_family="L2"
repeat_region	12605..12876	/rpt_family="L1"
repeat_region	12877..13185	/rpt_family="Alu"
misc_feature	13162..13185	/note="similar to Homo sapiens ERV1" (NID:gi1061813)"
repeat_region	13186..13221	/rpt_family="L1"
repeat_region	13223..13488	/rpt_family="L1"
repeat_region	13489..13750	/rpt_family="Alu"
repeat_region	13751..14391	/rpt_family="L1"
repeat_region	14332..14380	/rpt_family="MERL_type"
repeat_region	15690..15714	/rpt_family="(T)n"
repeat_region	15873..16172	/rpt_family="Alu"
repeat_region	16173..16300	/rpt_family="Alu"
repeat_region	16910..16942	/rpt_family="AT-rich"
repeat_region	17130..17499	/rpt_family="ERV1"
repeat_region	17303..17379	/rpt_family="A-rich"
repeat_region	17553..17621	/rpt_family="(CATATA)n"
repeat_region	17920..17970	/rpt_family="GA-rich"
repeat_region	18178..18658	/rpt_family="L1"
repeat_region	18644..18953	/rpt_family="L1"
repeat_region	19026..19155	/rpt_family="L2"
repeat_region	19061..19084	/rpt_family="AT-rich"
repeat_region	19167..19441	/rpt_family="L1"
repeat_region	19638..19749	/rpt_family="L1"
repeat_region	19850..20000	/rpt_family="L1"
repeat_region	20251..20438	/rpt_family="MER2_type"
repeat_region	21123..22815	/rpt_family="L1"
repeat_region	23367..23479	/rpt_family="MERL_type"
repeat_region	24536..24570	/rpt_family="(TCA)n"
repeat_region	24570..24849	/rpt_family="Alu"
repeat_region	25640..25700	/rpt_family="(TA)n"
repeat_region	26134..26688	/rpt_family="ERV1"
repeat_region	26944..27375	/rpt_family="MaLR"
repeat_region	27805..28356	/rpt_family="L2"
repeat_region	28532..28611	/rpt_family="L1"


```

|||||
102220 GGCCTCTGCTCCAGCAGTTGGGAGACCATATATATATGAGGAGTATATATGAGGAGG 102171
17 yGlyThrArgAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaA 34
102170 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 102121
34 laAlaGluAsnGluValHisPhePheAspTrpGluGluHisTyrSerHis 50
102120 CCCCCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 102071
51 GlyLeuGlyTrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLe 67
102070 GGGTTGGGTTGGTACATATATATATATATATATATATATATATATATATATATATAT 102021
67 uThrValGluLysThrProAlaTyrPheThrSerProLysValProGluA 84
102020 CACAGTGGAGAAATATATATATATATATATATATATATATATATATATATATATATAT 101971
84 rgValTyrSerMetAsnProSerIleArgLeuLeuLeuLeuLeuArgAsp 100
101970 GAGTCTACAGCATGAACCGGTCATCCCGCTGCTGCTCATCTCGGAGAC 101921
101 ProSerGluArgValLeuSerAspTyrThrGlnValPheTyrAsnHisMe 117
101920 CCGTCGGAGCGGCTATATATATATATATATATATATATATATATATATATATATATAT 101871
117 tGlnLysHisLysProTyrProSerIleGluGluPheLeuValArgAspG 134
101870 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 101821
134 lyArgLeuAsnValAspTyrLysAlaLeuAsnArgSerLeuTyrHisVal 150
101820 GCAGGCTCAATGTGGACTACAAGGCCCTCAACCGCAGCCTCTACCACGTG 101771
151 HisMetGluAsnTrpLeuArgPhePheProLeuArgHisIleHisIleVa 167
101770 CACATCGAAGCACTGGCTGGCTTTTTCGGCTGCGCCACATCCACATGT 101721
167 lAspGlyAspArgLeuIleArgAspProPheProGluIleGlnLysValG 184
101720 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 101671
184 luArgPheLeuLysLeuSerProGlnIleAsnAlaSerAsnPheTyrPhe 200
101670 AGAGGTTCTCTAAAGCTGTCGCCGACATCAATGCTTCCGAACCTCTACTTT 101621
201 AsnLysThrLysGlyPheTyrCysLeuAlaAspSerGlyAcqAspArgCy 217
101620 AACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 101571
217 sleuHisGluSerLysGlyArgAlaHisProGlnValAspProLysLeuL 234
101570 CTTACATGATGTCACAAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 101521
234 euAsnLysLeuHisGluTyrPheHisGluProAsnLysLysPhePheGlu 250
101520 TCAATAAAGTGCAGCAATATTTTCATGAGCCAAATGAAGTTCCTTCGAG 101471
101470 CTGTGTCGAGACATTTGACTGGGCAC 101444

```

seq_name: gb_ro:BC009133

seq_documentation_block:

LOCUS BC009133

1657 bp mRNA linear pop 12-JUL-2001

DEFINITION Mus musculus, heparan sulfate (glucosamine) 3-O-sulfotransferase 1,

clone MGC:11450 IMAGE:3155049, mRNA, complete cds.

ACCESSION BC009133

VERSION BC009133.1 GI:14318662

KEYWORDS MGC.

SOURCE house mouse.

ORGANISM

Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

Mus musculus

Eukaryota, Metazoa, Chordata, Vertebrata, Euteleostomi,

Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Mus,

1 (bases 1 to 1657)

Strausberg, R.

Direct Submission

Submitted (05-JUN-2001) National Institutes of Health, Mammalian

Gene Collection (MG), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC project URL: http://mgc.ncl.nih.gov

Contact: MGC help desk

Email: mgc-help@mail.nih.gov

Tissue Procurement: Lohar Hennighausen Ph.D., Chu-Xia Deng Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINI.)

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: http://www.hqsc.bcm.tmc.edu/cdna/

Contact: villalobosbcm.tmc.edu.

Villalobos, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,

A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,

Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/lini at: http://image.lini.gov

Series: IRAC Plate 16 Row: 5 Column: 17

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRFA 31 2618q70.

FEATURES

Location/Qualifiers

1..1657

/organism="Mus musculus"

/db_xref="LocusID:15476"

/db_xref="taxon:10090"

/clone="MGC:11450 IMAGE:3155049"

/tissue_type="Mammary tumor, Breast/fil: MMTV-Cre model. 10

months old, gross tissue."

/clone_lib="NCI_CGAP_Mam3"

/lab_host="DH10B"

/note="Vector: pCMV-SPORT6"

284..1219

/codon_start=-1

/product="heparan sulfate (glucosamine)

3-O-sulfotransferase 1"

/protein_id="AAH09133.1"

/db_xref="GI:14318663"

/translation="MTLLLGAVLLVAQPOLVHSHEAAAPGKLUQELLKVKVILPED

TGESTASNGSTQQLPOTLIIGVKGGSIFALLEMLSDHPVAAAEVHEFFUEEYISQ

GLAWYLCTMPFSSPHGLVFKTPVPTSPKVPPTTHSMNPTPLDLLIDKPSERVLSD

YTVQVLYNHQKHKPYPIEDLMRDGLINIDYKALNRSLYHAHMLNWLRFPLGHIHI

VDSIPRLIPFPFPIQKVPKPTKISPGINASNFYFNKTKGYCIKNSGKIDQCHESKGR

AHQVDKPLLDKLHEYPHEPNKPKFKLVGRTFQW"

BASE COUNT 398 a 439 c 393 g 427 t

ORIGIN

alignment_scores:

Quality: 1318.00 Length: 259

Ratio: 5.169 Gaps: 0

Percent Similarity: 98.456 Percent Identity: 92.278

alignment_block:

US-09-557-262-4_Copy_49_307 x BC009133

Align seg 1/1 to: BC009133 from: 1 to: 1657

1 GlySerAlaGlnLeuProGlnThrIleIleGlyValArgLysG1 17

|||||:|||||

440 GGTTCACACACACAGCAGTCGCCACAGCATCATCATTTGGGGTCCGCAAGG 489

|||||

17 yGlyThrArgAlaLeuGluMetLeuSerLeuHisProAspValAlaA 34

|||||

```

490 TGGTACCGGAGGCGCTGTATAGACATAGATAGATGATGATGATGATGATGATG 530
34  IAAAlaGluAsnGluValHisPhePheAspTrpGluGluHisTyrSerHis 50
|||||
540 CACCTCAAAACACGAGCTATCTCTTCTGACTGGGAGGACCATACAGCCAA 589
|||||
51 GlyLeuGlyTyrTyrLeuSerGlnMetProPheSerTrpProHisGlnLe 67
|||||
590 GCGCGCGCGCTGTATACCTACCTACCTACCTACCTACCTACCTACCTACCT 639
|||||
67  uthrValGluLysThrProAlaIlePheThrSerProLysValProGlu 84
|||||
640 CACCTGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 689
|||||
84  rGValTyrSerMetAsnProSerIleAlaGluLeuLeuIleLeuAsp 100
|||||
690 GAATTCACAGCATCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 739
|||||
101 ProSerGluArqValLeuSerAspTyrThrGlnValPheTyrAsnHisMe 117
|||||
740 CCATCAGACCGCGTCTGTCGACACACACACACACACACACACACACACCT 789
|||||
117 tGlnLysHisLysProTyrProSerIleGluGluPheLeuValArgAspG 134
|||||
790 TCAAGAAACAAAGAAATATATACCAATATAGACCTCTTAATGGGGACG 839
|||||
134 IyArgLeuAsnValAspTyrTyrAlaLeuAsnArgSerLeuTyrHisVal 150
|||||
840 GCGCGCTCAACCTGACCTACACAGCTCTTCAACCTGACCTGACCTGACCT 889
|||||
151 HisMetGluAsnTrpLeuArqPhePheProLeuArgHisIleHisIleVal 167
|||||
890 CACATGCTGCAACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 939
|||||
167 IAspGlyAspArgGluLeuArqAspProPhePheGluLeuGlnLysValG 184
|||||
940 GATATGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 989
|||||
184 IuArgPheLeuLysLysSerProGlnIleAsnAlaSerAsnProTyrPhe 200
|||||
990 AAAGATGCTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1039
|||||
201 AsnLysTrpLysGlyPheTyrCysLeuArqAspSerGlyArgAspArgC 217
|||||
1040 ACAAACCAACCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1089
|||||
217 HisHisGluSerLysGlyArgAlaHisProGlnValAspProLysLeuL 234
|||||
1090 CTATACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1139
|||||
234 euAsnLysLeuHisGlyTyrPheHisGluTrpAsnLysLysPheGlu 250
|||||
1140 TGCATAAACTCCAGCAATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1189
|||||
251 LeuValGlyArgThrPheAspTrpHis 259
|||||
1190 CTCGTGGGCAACAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1216
|||||
seq_name: gb_09-557-262-4_copy_49_307.p2n.rge
seq_documentation_block:
LOCUS      AF019385                      1685 bp    mRNA    linear    ROD 14-NOV-1997
DEFINITION Mus musculus heparan sulfate D-glucosaminyl 3-O-sulfotransferase-1
            precursor (3OST1) mRNA, alternatively spliced, complete cds.
ACCESSION  AF019385
VERSION    AF019385.1  GI:2618970
KEYWORDS   1 (bases 1 to 1685)
SOURCE     house mouse.
ORGANISM   Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE  1 (bases 1 to 1685)
AUTHORS   Shworak,N.W., Liu,J., Fritze,L.M.S., Schwartz,J.J., Zhang,L.,
            Loqeart,D. and Rosenberg,R.D.
            Molecular cloning and expression of mouse and human cDNAs encoding
            heparan sulfate D-glucosaminyl 3-O-sulfotransferase
            J. Biol. Chem. 272 (44), 28094-28019 (1997)
            98010647
            2 (bases 1 to 1685)
            Shworak,N.W., Liu,J., Fritze,L.M.S., Schwartz,J.J., Zhang,L.,
            Loqeart,D. and Rosenberg,R.D.
            Direct Submission
            Submitted (14-AUG-1997) Biology, MIT, 31 Ames Street, Cambridge, MA
            02139, USA
FEATURES             Location/Qualifiers
     source            1..1685
                        /organism="Mus musculus"
                        /strain="C3H/An"
                        /db_xref="taxon:10090"
                        /ccil_line="clonal; t. cell; line 17A clone 1A (described in
                        Shworak et. al., 1996, J. Biol. Chem. 271:27071-82)"
                        /note="poly(A)+ RNA isolated from postconfluent cultures"
     5'UTR             1..322
                        /gene="3OST1"
                        /note="alternatively spliced; this version represents a
                        Class 1 mRNA"
     gene              1..1685
     misc_feature      216..217
                        /gene="3OST1"
                        /note="splice variant site; many additional 3-OST-1 cDNAs
                        (Class 2 to Class 9) contain extra 5' UTR exons between
                        these nucleotides (extra 86 to 1629 bp)"
     sig_peptide        323..382
                        /gene="3OST1"
                        323..1258
                        /gene="3OST1"
                        /function="rate limiting enzyme for the synthesis of
                        antioquant heparan sulfate"
                        /note="heparan sulfate sulfotransferase, 3 OST-1;
                        interluminal Golgi resident protein (retention mechanism
                        unknown)"
                        /codon_start=1
                        /product="heparan sulfate D-glucosaminyl
                        3-O-sulfotransferase-1 precursor"
                        /protein_id="AA084387.1"
                        /db_xref="GI:2618971"
     mat_peptide        383..1245
                        /gene="3OST1"
                        /note="confirmed by N-terminal analysis"
                        /evidence=experimental
                        /product="heparan sulfate D-glucosaminyl
                        3-O-sulfotransferase-1"
                        476..484
                        /gene="3OST1"
                        /note="encodes N-linked glycosylation sites"
                        /evidence=experimental
                        494..1255
                        /gene="3OST1"
                        /note="encodes presumptive sulfotransferase catalytic
                        domain"
                        908..916
                        /gene="3OST1"
                        /note="encodes N-linked glycosylation sites"
                        /evidence=experimental
                        1058..1066
                        /gene="3OST1"
                        /note="encodes N-linked glycosylation sites"
                        /evidence=experimental
                        1079..1087

```

```

/misc_structure /gene="30stl"
/notes="encodes N-linked glycosylation sites"
/evidence="experimental"
1100..1129
/gene="30stl"
/notes="encodes cysteine-bridged peptide loop"
/evidence="experimental"
1259..1285
/gene="30stl"
1532..1537
/gene="30stl"
1536
/gene="30stl"
1586..1591
/gene="30stl"
1590..1595
/gene="30stl"
1605
/gene="30stl"
1610
/gene="30stl"
1611..1615
/gene="30stl"
1629
/gene="30stl"
1667..1672
/gene="30stl"
1685
/gene="30stl"
BASE COUNT 422 a 426 c 382 g 455 t
ORIGIN

```

```

alignment_scores:
  Quality: 1318.00      Length: 259
  Ratio: 5.169          Gaps: 0
  Percent Similarity: 98.456  Percent Identity: 92.278

alignment_block:
US-09-557-262-4_copy_49_307 x AF019385 ..
Align seq 1/1 to: AF019385 from: 1 to: 1645

1 GlySerAlaGlnGluLeuProGlnThrIleIleGlyValArqLysG1 17
|||||
479 GGTTCACACACACAGCTGCCACACACACACACACACACACACACAGG 528
|||||
17 yGlyThrArqAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaA 34
|||||
529 TGGTACCGAGACCTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 578
|||||
34 IaAlaGlnAsnGluValHisphoAspTrpGluGluHisTyrSerHis 50
|||||
579 CAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 528
|||||
51 GlyLeuGlyTrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLe 57
|||||
629 GGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCG 678
|||||
67 uThrValGluLysThrProAlaTyrPheThrSerProLysValProGlnA 84
|||||
679 CACCTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 728
|||||
84 rGValTyrSerMetAsnProSerIleArgGluLeuLeuIleLeuArqAsp 100
|||||
729 GAATGCAATATATATATATATATATATATATATATATATATATATATAT 778
|||||
101 ProSerGluArqValLeuSerAspTyrThrGlnValPheTyrAsnHisMe 117
|||||
779 CATACAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 828
|||||
117 tGlnLysHisTyrProTyrProSerIleGluPheLeuValArqAspG 134
|||||

```

```

829 TCAGAAAGAAAGAGGCTATCCAGAGATATAGAGAGATCTTAATGAGGAG 878
134 IyArqLeuAsnValAspTyrLysAlaLeuAsnArgSerLeuTyrHisVal 150
|||||
879 GTCGGCTCAAGCTGGACACAGAGCTTCACAAAGCCAGCCGTAGCAAGCA 928
|||||
151 HisMetGlnAsnTrpLeuArqPhePheProLeuArqHisIleHisIleVa 167
|||||
929 CACATGCTCAACAGGCTGCGGTTCGGGTTCGGGTTCGGGTTCGGGTTCG 978
|||||
167 AspGlyAspArgLeuIleArgAspProPheProGluIleGlnLysValG 184
|||||
978 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1028
|||||
184 IuArqPheLeuLysLeuSerProGlnIleAsnAlaSerAsnPheTyrPhe 200
|||||
1029 AAGATTCCTCAAGCTTCTCCACAGATCAACGCTCGAACCTCTACCTT 1078
|||||
201 AsnLysThrLysGlyPheTyrCysLeuArqAspSerGlyArqAspArgG 217
|||||
1079 AAAAAAACAAGGCTTCTACTGCTTGGCGCACAGTGGCAAGGACCGCTG 1128
|||||
217 SleuHisGluSerLysGlyArgAlaHisProGlnValAspProLysLeuL 234
|||||
1129 CTTACACAGAGTCAAAAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAT 1178
|||||
234 cuAsnLysHisGlnSerTyrPheHisGluProAsnLysLysPhePheGlu 250
|||||
1179 TTGATAAACTCCAGCAATACITTCATGAGCCAAATACAAAATTTTTCAG 1228
|||||
251 LeuValGlyArgThrPheAspTrpHis 259
|||||
1229 CTCGTGGCGAGAACATTCGACTGGCAC 1255
|||||
seq_name: gb_htg.AC084416_3
seq_documentation_block:
WPCOMMENT
Sequence split into 6 fragments locus AC084416 Accession AC084416
Fragment Name Begin End
AC084416_0 1 110000
AC084416_1 499901 510000
AC084416_2 200001 310000
AC084416_3 300001 410000
AC084416_4 400001 510000
AC084416_5 500001 595289
Continuation (4 of 6) of AC084416 from base 300001 (AC084416 Mus musculus clone rp233)

alignment_scores:
  Quality: 1318.00      Length: 259
  Ratio: 5.169          Gaps: 0
  Percent Similarity: 98.456  Percent Identity: 92.278

alignment_block:
US-09-557-262-4_copy_49_307 x AC084416_3/rev ..
Align seq 1/1 to: AC084416_3 from: 1 to: 110000

1 GlySerAlaGlnGluLeuProGlnThrIleIleGlyValArqLysG1 17
|||||
77721 GATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 77672
|||||
17 yGlyThrArqAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaA 34
|||||
77671 TGGTACCGAGACCTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 77622
|||||
34 IaAlaGlnAsnGluValHisphoAspTrpGluGluHisTyrSerHis 50
|||||
77621 CAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 77572
|||||
51 GlyLeuGlyTrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLe 67
|||||

```



```

1  GlySerAlaGlnGlnLeuProGlnThrIleIleIleIleIleValArgLysG1 17
|||||...|||||
112027 GCCTCCACACAGCAGCTGCCACAGACCATCATTCATTCGGGTGCCCAAGGC 112076

17  yGlyThrArgAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaA 34
|||||...|||||
112077 TGGTACCCGAGGATCTCTAGAGATCTTAGCTGATCTCTGATCTGCTG 112126

34  IaAlaGluAsnGluValHisPheAspTrpGluGluHisTyrSerHis 50
|||||...|||||
112127 CAGCTGAAACGAGGCTGCTATCTTCTTACTGAGAGAGCATTAACAGCAA 112176

51  GlyLeuGlyTrpTrpLeuSerGlnMetProPheSerTrpProHisGlnLe 67
|||||...|||||
112177 GGCCTGGGCTGCTACCTACCCACAGATCCCTTCCTCCCTCCACCACT 112226

67  uThrValGluLysThrProAlaTyrPheThrSerProLysValProGluA 84
|||||...|||||
112227 CAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 112276

84  rGValTyrSerMetAsnProSerIleArgLeuLeuLeuLeuLeuArgAsp 100
|||||...|||||
112277 GAATCCACAGCATGAACCCACCATCCGCTGCTTATCTCAGAGAC 112326

101 ProSerGluArgValLeuSerAspTyrThrGlnValPheTyrAsnHisMe 117
|||||...|||||
112327 CCATCAGAGCGGCTGCTGCTGCTACACACAGGCTGTGTGACACCACT 112376

117  tGlnLysHisLysProTyrProSerIleGluGluPheLeuValArgAspG 134
|||||...|||||
112377 TCAGAGACAGAGCTTATCAGCTATCAGCTATCAGCTATCAGCTATC 112426

134  lYArgIleAsnValAspTyrLysAlaLeuAsnArgSerLeuTyrHisVal 150
|||||...|||||
112427 GTGGGTGAACCTGAGATACAGAGCTTATCAGCTATCAGCTATCAG 112476

151 HisMetGlnAsnTrpLeuArgPhePheProLeuArgHisIleHisIleVa 167
|||||...|||||
112477 CACATGCTGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 112526

167  lAspGlyAspArgLeuIleArgAspProPheProGluIleGlnLysValG 184
|||||...|||||
112527 GGATGGGAGCGGCTCATCAGAGAGCTTTCCTGAGATCCAGAGGTGG 112576

184  luArgPheLeuLysLeuSerProGlnIleAsnAlaSerAsnPheTyrPhe 200
|||||...|||||
112577 AAAGATTCCTGAAGCTTTCTCCACAGATCAAGGCTGCTGCTGCTG 112626

201 AsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyArgAspArgCy 217
|||||...|||||
112627 AACAAACCAAGCGCTTCTACTGCTGCTGCTGCTGCTGCTGCTGCTG 112676

217  sLeuHisGluSerLysGlyArgAlaHisProGlnValAspProLysLeuL 234
|||||...|||||
112677 CTACACAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 112726

234  euAsnLysLeuHisGluTyrPheHisGluProAsnLysLysPhePheGlu 250
|||||...|||||
112727 TTGATAAAGCTGACGAAATCTTCTATAGCCAAATTAAGAAATTTTCA 112776

251 LeuValGlyArgThrPheAspTrpHis 259
|||||...|||||
112777 CTCGTGGGAGAACATTCGACTGGCAC 112803

seq_name: qb_ro:AF177430
seq_documentation_block:
LOCUS AF177430 936 bp mRNA linear ROD 10-MAY-2001
DEFINITION Rattus norvegicus 3-O-sulfotransferase mRNA, complete cds.
ACCESSION AF177430
VERSION AF177430.1 GI:9957243
KEYWORDS

```

```

SOURCE
ORGANISM
Norway rat.
Rattus norvegicus
Pukaryota; Metazoa; Chordata, Craniata, Vertebrata, Euteleostomi;
Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae;
Rattus.
REFERENCE
1 (bases 1 to 936)
Li, Z. Y., Hirayoshi, K. and Suzuki, Y.
Expression of N-deacetylase/sulfotransferase and
3-O-sulfotransferase in rat alveolar type II cells
Am. J. Physiol. 279 (2), L292-L301 (2000)
JOURNAL
2 (bases 1 to 936)
Li, Z. Y., Hirayoshi, K. and Suzuki, Y.
Direct Submission
TITLE
Submitted for Frontiers Medical Sciences, 53 Shogoin Kawahara-cho,
Sakyo-ku, Kyoto 606 8507, Japan
FEATURES
location/Qualifiers
1..936
/organism="Rattus norvegicus"
/surstrain="wistar"
/db_xref="taxon:10116"
/cell_type="type II epithelium"
1..936
/note="3' 5' UTR"
/codon_start=1
/product="3-O-sulfotransferase"
/protein_id="AAC09283.1"
/db_xref="GI:9957244"
/translation="MTLLLLGAVLLVAQVLPVSPHAAAGHCLKQGLIKKVLIIIPED
TGEAATNGSTQQLPOTIIIGVRKGTGTRALLEMLSDPDAANEVHFFDWEHYSQ
GLWYLTQMPFSFQHTVEKTPAYTPKPEVHHSNPTIRLLILRDPSEVLSD
YTGVLYNHILQKHKPYFIEDILMPDGRNLNVDYKALNPSLYHAIHMLMWLPFPLGHIH
VDGRPIIDPPPELQKVERFLKSPQINASNPYKTKGFYCLRDGSKDRCHHSKGR
AHPQVDPKLLDKLHEVFRPNKFTKLKVFDMH"
BASE COUNT 224 a 292 c 224 g 196 t
ORIGIN

alignment_scores:
Quality: 1304/300 Length: 259
Ratio: 5.133 Gaps: 0
Percent similarity: 98.456 Percent identity: 91.892

alignment_block:
US-09-557-262-4_COPY_49_307 x AF177430 ..
Align seg 1/1 to: AF177430 from: 1 to: 936

1  GlySerAlaGlnGlnLeuProGlnThrIleIleIleIleIleValArgLysG1 17
|||||...|||||
157  GCTTCCACACAGCAGCTGCCACAGACCATCATTCATTCGGGTGCCCAAGGC 206

17  yGlyThrArgAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaA 34
|||||...|||||
207  TGGTACCCGAGGATCTCTAGAGATCTTAGCTGATCTCTGATCTGCTG 256

34  IaAlaGluAsnGluValHisPheAspTrpGluGluHisTyrSerHis 50
|||||...|||||
257  CAGCTGAAACGAGGCTGCTATCTTCTTACTGAGAGAGCATTAACAGCAA 406

51  GlyLeuGlyTrpTrpLeuSerGlnMetProPheSerTrpProHisGlnLe 67
|||||...|||||
307  GGCCTGGGCTGCTACCTACCCACAGATCCCTTCCTCCCTCCACCACT 356

67  uThrValGluLysThrProAlaTyrPheThrSerProLysValProGluA 84
|||||...|||||
357  CAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 406

84  rGValTyrSerMetAsnProSerIleArgLeuLeuLeuLeuLeuArgAsp 100
|||||...|||||
407  GAATCCACAGCATGAACCCACCATCCGCTGCTTATCTCAGAGAC 456

101 ProSerGluArgValLeuSerAspTyrThrGlnValPheTyrAsnHisMe 117

```

```

|||||
457 CCATCCGAACGTCGCTATCCCACTACACCCAGCTGTCTACAAACCCCT 506
|||||
117 TGLNLSHLSLysProTyrProSerIleGluGluPheLeuValArqAspG 134
|||||
507 TCAGAAACACAAATCTACACCACTATGACAGCTCTTAAGAGGAGAG 556
|||||
134 TATGluLeuValAspTyrTyrLysAlaLeuAsnArqSerLeuTyrHisVal 150
|||||
557 GAGGAGTCAGATGAGTATCAAGAGTCTCAAGAGGCAATCTATATG 606
|||||
151 HisMetGlnAsnTyrPheArqPhePheProLeuArgHisIleHisIleVal 167
|||||
607 CACATGCTGCAACTGCTGCGGCTTTTTCCTACATGAGGAGGAGGAG 656
|||||
167 TAspGlyAspArqLeuIleArqAspProPheProGluIleGlnLysValG 184
|||||
657 GCAATGCGGACCGTTTCAACACAGACAGCTTTTCCGACAGTCCAGAA 706
|||||
184 TArqPheLeuLysLeuSerProGlnIleAsnAlaSerAsnPhetPhe 200
|||||
707 AAGGTTCTCAAGCTGTCTCAACAGATCAAGAGGCTCGAACTTCTACT 756
|||||
201 AsnLysThrTyrSerTyrPheTyrCysLeuArqAspSerClyArqAspArg 217
|||||
757 AACAAACCAAGGCTTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 806
|||||
217 sLeuIleGlnSerLysGlyArqAlaHisProGlnValAspProLysLeu 234
|||||
807 CTATACACAGCTCAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 856
|||||
234 cuAsnLysLeuHisGluTyrPheHisGluProAsnLysLysPhePheGlu 250
|||||
857 TTGATAAATGAGGATATTTTGTGAGGCAATAAAGAAATTTTTCAG 906
|||||
251 LeuValGlyArqThrPheAspTyrHis 259
|||||
907 CTCGTGGGACAGCAATTCGACTGGCAC 933
|||||
seq_name: qb_pat:AX317978
seq_documentation_block:
LOCUS AX317978 2115 bp DNA linear PAT 14-DEC-2001
DEFINITION Sequence 18 from Patent WO0190334.
ACCESSION AX317978
VERSION AX317978.1 GI:17900788
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (sites)
Yue, H., Sanjanwala, M.S., Baughn, M.R., Gandhi, A.R., Ring, H.Z.,
Ellrott, V., Walla, N.K., Yang, J., Khan, F.A., Ramkumar, J., Tang, Y.T.,
Hafalia, A., Lal, P., Nguyen, D.H., Yao, M.C., Lee, E.A.,
Tribouley, C.M., Patterson, C., Lu, Y., Ding, L., Bruns, C.M. and
Kearney, L.
JOURNAL Patent: WO 0190334-A 18 29-NOV-2001;
Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
1..2115
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 592468/CB1"
BASE COUNT 651 a 408 c 438 g 618 t
ORIGIN

```

alignment_scores:

Quality: 772.50 Length: 258
Ratio: 3.527 Gaps: 2
Percent Similarity: 84.884 Percent Identity: 53.101

alignment_block:

US-09-557-262-4_COPY_49_307 x AX317978

Align seq 1/1 to: AX317978 from: 1 to: 2115

```

4 ATGGLLeuPheGlnThrIleIleIleGlyValArqLysGlyGlyThrAl 20
|||||
438 CAGCAGCTCCCAAGGACATTAATCATTTGGGGTCAGGAAAGCAGGACAA 487
|||||
20 qAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaAlaIleGlu 47
|||||
488 GCGCTCTGCTTCAAAATGCTGAACTTACATCCGCGAGTAGTCAAAAGC 547
|||||
37 SngGluValHisPhePheAspIleIleGluHisTyrSerHisGlyLeuGly 53
|||||
538 AAGAAATCCACTTTTTCATTAATGATGAGAAATTAATGTAAGGCTAT 587
|||||
54 TrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLeuThrValG 70
|||||
588 TGGTATAGCAAAAGAGATCCCTTTTTCCTACCCCTCAGCAAAATCACA 637
|||||
70 uLysThrProAlaTyrPheThrSerProLysValProGluArqValTyr 87
|||||
638 AAGAGGCTCAACATATATATATATATATATATATATATATATATAT 687
|||||
87 erMetAsnProSerIleArqLeuLeuIleLeuArqAspProSerGlu 103
|||||
688 AATGAACTATCATATCAATCAATGTTGTAATATATATATATATATAT 747
|||||
104 ArqValLeuSerAspTyrThrGlnValPheTyrAsnHisMetGlnLysH 120
|||||
738 AGAGTATATATATATATATATATATATATATATATATATATATAT 787
|||||
120 sLysProTyrProSerIleGluGluPheLeuValArqAspGly..... 135
|||||
788 CAAACTTATATCAAGTTTCAGAGCTGGCCATAGACCCCTAATACATGG 837
|||||
135 rGluAsnValAspTyrTyrLysAlaLeuAsnArqSerLeuTyrHisVal 151
|||||
838 AAGTGAACACAAATAACAAAGCAGTAAGAAACAGATATATATATATAT 887
|||||
152 MetGlnAsnTrpLeuArqPhePheProLeuArqHisIleHisIleVal 168
|||||
888 CTGCAAAAGGTCGCTGAAATACITTCCTCAATTTGAGCAATTTCAATG 947
|||||
168 pGlyAspArqLeuIleArqAspProPheProGluIleGlnLysValGlu 185
|||||
938 TGGAGATCGCTTCATCAGCAACCTCTGCTCAACCTTCAGCTCGTGAG 987
|||||
185 rGluLeuLysLeuSerProGlnIleAsnAlaSerAsnPhetTyrHis 201
|||||
988 AGTTCTTAAATCTGCTCTCAAGGATAAGTCAATCAATTAATTAATTA 1037
|||||
202 LysThrLysGlyPheTyrCysLeuArq.....AspSerClyArqAspArg 217
|||||
1038 GCTACCAAGAGGTTTACTGCTTGGCGTTTAAATATATATCTTTAAAT 1087
|||||
217 sLeuHisGluSerLysGlyArqAlaHisProGluValAspProLysLeu 234
|||||
1088 CCTGGCGGCGAGCAAGGCGGCTTCATCCAGAGGTGGACCCCTCTGTCA 1137
|||||
234 cuAsnLysLeuHisGluTyrPheHisGluProAsnLysLysPhePheGlu 250
|||||
1138 TTAATAAAATGGCAATTCCTTTCATCTTTTAAATCAAAATTTTTCAG 1187
|||||
251 LeuValGlyArqThrPheAspTrp 258
|||||
1188 ATCACTGGGAGGACATTTGAACTGG 1211
|||||
seq_name: qb_pat:AL355498
seq_documentation_block:
LOCUS AL355498 165549 bp DNA linear PRI 10 JAN-2001

```


Human DNA sequence from clone RP11-112L15 on chromosome 6, complete sequence.

ACCESSION AI355498
VERSION AI355498.10 GI:112141339
KEYWORDS HGC
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Placentalia; Primates, Catarrhini, Hominoidea; Homo.
1 (bases 1 to 165549)
Hall, R.
Direct Submission
Submitted (09-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, UK
CB10 1SA, UK E-mail enquiries: hgmquery@sanger.ac.uk
requests: hgmquery@sanger.ac.uk
On Jan 12, 2001 this sequence version replaced gi:11544487.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EM, EMBL; SW, SWISSPROT; Tr, TrEMBL; Wp, WormPEP. Information on the WormPEP database can be found at http://www.sanger.ac.uk/projects/c_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>
RP11-112L15 is from the library RP11-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/ba/pap/home.htm>
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP11 112L15. The true left end of clone RP11-44K2 is at 55185 in this sequence.

FEATURES
source
Location/Qualifiers
1..165548
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-112L15"
/clone.lib="RP11-11.1"
repeat_region 256..573
/note="AluX repeat. matches 1..312 of consensus"
repeat_region 1685..1988
/note="AluJb repeat: matches 1..298 of consensus"
repeat_region 2013..2318
/note="AluSg repeat: matches 1..305 of consensus"
repeat_region 2776..2807
/note="16 copies 2 mer at 84% conserved"
repeat_region 2864..3163
/note="AluX repeat: matches 1..293 of consensus"
repeat_region 4069..4240
/note="MIP repeat: matches 105..262 of consensus"
repeat_region 4737..4831
/note="12 repeat: matches 2108..2710 of consensus"
repeat_region 5020..5364
/note="11pA6 repeat: matches 5798..6143 of consensus"
misc_feature 5187..5197
/note="1372 bases of IS186 transposon (X03123) removed here. This sequence represents the duplicated flanking sequence of the IS186."
repeat_region 5646..6640
/note="11p4 repeat: matches 4784..5781 of consensus"
repeat_region 7170..7271
/note="34 copies 3 mer at 61% conserved"


```

/misc_feature /gene-"3OST2"
313..315 /note-"putative O-glycosylation site"
/gene-"3OST2"
/misc_feature /note-"putative O-glycosylation site"
322..324
/gene-"3OST2"
/misc_feature /note-"putative O-glycosylation site"
331..333
/gene-"3OST2"
/misc_feature /note-"putative O-glycosylation site"
367..375
/gene-"3OST2"
/misc_feature /note-"putative N-glycosylation site"
400..417
/gene-"3OST2"
424..462 /note-"sulfolransferase domain, unclassified site"
/gene-"3OST2"
/misc_feature /note-"putative beta-strand/p-loop/alpha-helix motif;
binds 5'-phosphate of 3'-phosphoadenosine 5'-phosphate,
PAPS; binding site"
649..657
/gene-"3OST2"
/misc_feature /note-"putative N-glycosylation site"
667..714
/gene-"3OST2"
/misc_feature /note-"putative sheet-loop-helix structure for binding the
3'-phosphate of PAPS; binding site"
730..732
/gene-"3OST2"
/misc_feature /note-"putative O-glycosylation site"
775..783
/gene-"3OST2"
/variation /note-"putative N glycosylation site"
804
/gene-"3OST2"
/misc_feature /note-"3OST2-2 allele, silent wobble position mutation"
823..852
/replacement
/gene-"3OST2"
/misc_feature /note-"putative structure for binding 5'-sulfate of PAPS;
binding site"
986..996
/gene-"3OST2"
/misc_feature /note-"putative N glycosylation site"
1009..1047
/gene-"3OST2"
/misc_feature /note-"putative cysteine bridged peptide loop; unclassified
site"
1102..1110
/gene-"3OST2"
/misc_feature /note-"putative N-glycosylation site"
1177..1190
/gene-"3OST2"
/variation /note-"3OST2"
1249
/gene-"3OST2"
/misc_feature /note-"3OST2-2 allele"
1350
/replacement
/gene-"3OST2"
/variation /note-"3OST2-2 allele"
1350
/replacement
1507
/gene-"3OST2"
/misc_feature /note-"3OST2-2 allele"
1743..1748
/gene-"3OST2"
/polyA_signal /note-"putative"
1925..1930
/gene-"3OST2"
/polyA_site 1950

```

```

BASE COUNT      441 a      603 c      490 g      434 t
ORIGIN

alignment_scores:
  Quality: 598.00      Length: 261
  Ratio: 2.990      Gaps: 4
  Percent Similarity: 76.628      Percent Identity: 44.295

alignment_block:
US-09-557-262-4_COPY_49_307 x AF105374
Align seq 1/1 to: AF105374 from: 1 to: 1968

      4  GlnGlnLeuProGlnThrIleIleIleGlyValArgLysGlyGlyThr 20
      406  AAGCGTTCGGGCAAGAGGCGTATTGTTGGGCTGAAGAGGGGGCGACGG 455
      20  qAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaAlaAlaAla 37
      456  GATTCGTTCGAGTTATCCAGTACACCGGACGTGGCGCTTGGGCA 505
      37  ssGluValHisPhePheAspTrpGluGluHisTyrSerHisGlyLeuGly 53
      506  CGAAGCCCACTTCTTGAC      AGGAAATAGGCGCGGCGCTGAT 549
      54  TrpTyrLeuSerGlnMetProHisSerTrpProHisGlnLeuThrVal 70
      550  TGGTATAGAGGCTGATAGCGAGAGAGGCTGAGAGAGAAAGATCAGTGG 599
      70  uLysThrProAlaTyrPheThrSerProLysValProGluArgValTyrS 87
      600  CAAGACGCGGAGTACTTTGTCACACAGAGAGGCTCCGACGAGACATCA 649
      87  erMetAsnProSerIleArgLeuLeuLeuLeuArgAspProSerGlu 103
      655  ATATCTTCGAGAAAGAGAGATGATGTTGTTGAGGAGAACTGTGAG 699
      104  ArgValLeuSerAspTyrThrGluValPheTyrAsnHisMetGlnLysHis 120
      700  GGTGACATCTGTATTACAGCAGACA      CTTCCCAAGAA 737
      120  sLysProTyrProSerIleGluGluPheLeuValArgAsp      G 134
      738  GCGGCAATACCGGAGCTTGAAGGCGCTCTCTCCGCAAGCGGACGCTGG 787
      134  lyArgLeuAsnValAspTyrLysAlaLeuAsnArgSerLeuTyrHisVal 150
      788  GGTGTGAGAGAGTGTGTGAAGAAATGGATATGGCATGATGAGTGGCTG 847
      151  HisMetGlnAsnTrpGluArgPheProLeuArgHisIleHisIleVal 167
      838  CAGTGGAGAGAGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 887
      167  lAspGlyAspArgLeuLeuArgAspProPheProGluIleGlnLysVal 184
      888  CAGTGGAGAGAGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 937
      184  luArgPheLeuLysLeuSerProGlnIleAsnAlaSerAsnPheTyrPhe 200
      938  AGGATTTCTGGAGATTAGAGATTCAGAGAGAGAGAGAGAGAGAGAGAG 987
      201  AsnLysThrLysGlyPheTyrCysLeuArg      AspSerGlyAla 214
      988  AACAGAGCAAGAGATTCTGTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1037
      214  gAspArgCysLeuHisGluSerLysGlyArgAlaHisProGlnValAsp 231
      1038  GCGTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1087
      231  roLysLeuLeuAsnLysLeuHisGluTyrPheHisGluProAsnLys 247

```


[illegible]

100



Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-557-262-4_copy_246_272 x ABA50047/rev

Align seq 1/1 to reverse of: ABA50047 from: 1 to: 924

```

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyAla 17
|||||
170 TTCTACTTTAACAACCAACGCGCTTTTACACGCTCGGCGACACGCGCG 121
|||||
17 gAspArgCysLeuHisGluSerLysGlyArg 27
|||||
120 GGAAGAGTGTATTATATGATTTTAAAGAGTGG 90

```

seq_name /SINSS/gcgdata/gene-seq/gene-seq-emb1/NA2001A.DAT.ABA67978

seq_documentation_block:

ID ABA67978 standard; DNA; 924 BP.

XX AC ABA67978;

XX DT 01-FEB-2002 (first entry)

XX DE Human foetal liver single exon nucleic acid probe #16283

XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX OS Homo sapiens.

XX PN W0200157277-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00666.

XX PR 04-FEB-2000; 2000US 0180312.

XX PR 26-MAY-2000; 2000US 0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US 0632366.

XX PR 21-SEP-2000; 2000US-0234487.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000US 0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX XX WPI; 2001-483447/52.

XX PT Human genome derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human fetal liver -

XX PS Claim 4, SEQ ID NO 16283, 637pp + sequence listing, English.

XX CC The invention relates to a single exon nucleic acid probe for

XX CC measuring human gene expression in a sample derived from human

XX CC liver. The single exon nucleic acid probes may be used for predicting,

XX CC measuring and displaying gene expression in samples derived from human

XX CC fetal liver. The present sequence is a single exon nucleic acid

XX CC probe of the invention.

XX CC Note: The sequence data for this patent did not form part of the

XX CC printed specification, but was obtained in electronic format directly

XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 924 BP; 169 A; 252 C; 310 G; 193 T; 0 other;

alignment_scores:

Quality: 155.00 Length: 27
Ratio: 5.741 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-557-262-4_copy_246_272 x ABA67978/rev

Align seq 1/1 to reverse of: ABA67978 from: 1 to: 924

```

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyAla 17
|||||
170 TTCTACTTTAACAACCAACGCGCTTTTACACGCTCGGCGACACGCGCG 121
|||||
17 gAspArgCysLeuHisGluSerLysGlyArg 27
|||||
120 GGAAGAGTGTATTATATGATTTTAAAGAGTGG 90

```

seq_name /SINSS/gcgdata/gene-seq/gene-seq-emb1/NA2001A.DAT.ABA35017

seq_documentation_block:

ID ABA35017 standard; DNA; 924 BP.

XX AC ABA35017;

XX DT 23-JAN-2002 (first entry)

XX DE Probe #13483 for gene expression analysis in human heart cell sample.

XX KW Human; gene expression; heart; microarray; vascular system; probe;

XX KW cardiovascular disease; hypertension; cardiac arrhythmia;

XX KW congenital heart disease; ss.

XX OS Homo sapiens.

XX XX W0200157274-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00666.

XX PR 04-FEB-2000; 2000US 0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US 0632366.

XX PR 21-SEP-2000; 2000US-0234487.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000US 0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX XX WPI; 2001-488893/53.

XX PT Single exon nucleic acid probes for analyzing gene expression in human

XX PT hearts -

XX PS Claim 4, SEQ ID NO 13483, 536pp; English.

XX CC The present invention relates to single exon nucleic acid probes for

XX CC measuring human gene expression in a sample derived from human heart. The

XX CC present sequence is one such probe. The probes may be used for

XX CC predicting, measuring and displaying gene expression in samples derived

XX CC from the human heart via microarrays. By measuring gene expression, the

XX CC probes are useful for predicting, diagnosing, grading, staging,

XX CC monitoring and prognosing diseases of the human heart and vascular system

XX CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and

XX CC congenital heart disease.

XX CC Note: The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 924 BP; 169 A; 252 C; 310 G; 193 T; 0 other;

alignment_scores:

Quality: 155.00 Length: 27

Ratio: 5.741
Percent Similarity: 100.000 Percent Identity: 100.000
Caps: 0

alignment_block:

US-09-557-262-4_COPY_246_272 x ABA35017/rev

Align seg 1/1 to reverse of ABA35017 from: 1 to: 924

```

1 PheTyfHeAsnAspLysThrLysAlaPheTyfCysLeuArgAspSerGlyAr 17
|||||
170 TTTACTTTTACAAACCAAGGAGCTTTTACTGCTGTGAGGGAAGAGAGAG 121
|||||
17 GAspArgCysLeuHisGluSerLysGlyArg 27
|||||
120 GAGAGAGTGTATTATATATATATATATATATATATATATATATATATAT 90

```

seq_name: /SIDS5/gcadata/geneseq/geneseq-emb1/NA2001A.DAT:AAK16371

seq_documentation_block:

ID: AAK16371 standard; DNA: 924 BP.

XX AC AAK16371;

XX DT 06-NOV-2001 (first entry)

XX DE Human brain expressed single exon probe SEQ ID NO: 16362.

XX KW Human; brain expressed exon; gene expression analysis; probe;

XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

XX KW epilepsy; cancer; ss.

XX OS Homo sapiens.

XX PN WC200157276-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-0500667

XX PP 04-FER-2000; 2000US-0180312.

XX PP 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0609408

XX PR 03-AUG-2000; 2000US-0642466

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000US-0024263

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-483446/52.

XX PT Single exon nucleic acid probes for analyzing gene expression in human

XX PT brains

XX XX

XX Example 4; SEQ ID NO: 1562; 650pp + Sequence Listing, English.

XX CC The present invention provides a number of single exon nucleic acid

XX CC probes which are derived from genomic sequences expressed in the human

XX CC brain. They can be used to measure gene expression in brain cell samples,

XX CC which may enable the diagnosis and improved treatment of nervous system

XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

XX CC epilepsy and cancers. The present sequence is one of the probes of the

XX CC invention.

XX SQ Sequence 924 BP; 169 A; 252 C; 310 G; 193 T; 0 other;

alignment_scores:

Quality: 155.00

Ratio: 5.741

Percent Similarity: 100.000

Length: 27

Caps: 0

Percent Identity: 100.000

alignment_block:

US-09-557-262-4_COPY_246_272 x AAK16371/rev

Align seg 1/1 to reverse of AAK16371 from: 1 to: 924

```

1 PheTyfHeAsnAspLysThrLysAlaPheTyfCysLeuArgAspSerGlyAr 17
|||||
170 TTTACTTTTACAAACCAAGGAGCTTTTACTGCTGTGAGGGAAGAGAGAG 121
|||||
17 GAspArgCysLeuHisGluSerLysGlyArg 27
|||||
120 GAGAGAGTGTATTATATATATATATATATATATATATATATATATATAT 90

```

seq_name: /SIDS5/gcadata/geneseq/geneseq-emb1/NA2001A.DAT:AAK16371

seq_documentation_block:

ID: AAK16371 standard; DNA: 924 BP.

XX AC AAK16371;

XX DT 06-NOV-2001 (first entry)

XX DE Human bone marrow expressed single exon probe SEQ ID NO: 16679.

XX KW Human; bone marrow expressed exon; gene expression analysis; probe;

XX KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX OS Homo sapiens.

XX PN WC200157276-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-0500668.

XX PP 04-FER-2000; 2000US-0180312.

XX PP 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0609408

XX PR 04-AUG-2000; 2000US-0642466

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000US-0024263

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488900/53.

XX PT Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human bone marrow

XX XX

XX Example 4; SEQ ID NO: 16679; 658pp + Sequence Listing, English.

XX CC The present invention provides a number of single exon nucleic acid

XX CC probes which are derived from genomic sequences expressed in the human

XX CC bone marrow. They can be used to measure gene expression in bone marrow

XX CC samples, which may enable the improved diagnosis and treatment of cancers

XX CC such as lymphoma, leukaemia and myeloma. The present sequence is one of

XX CC the probes of the invention.

XX SQ Sequence 924 BP; 169 A; 252 C; 310 G; 193 T; 0 other;

alignment_scores:

Quality: 155.00

Ratio: 5.741

Percent Similarity: 100.000

Length: 27

Caps: 0

Percent Identity: 100.000

alignment_block:

US-09-557-262-4_COPY_246_272 x AAK16371/rev

Align seq 1/1 to reverse of: AAK42122 from: 1 to: 924

1 PheYrPheAsnYsThrYsGlyPheCysLeuAraAspSerGlyAr 17
 |||||
 170 TTCTACTTTAAACAAACCAAGGGCTTTTACTGCTCGGACAGCGCGG 121

17 qAspAraCysLeuIleGlySerLysGlyArq 27
 |||||
 120 GACCGCGTGTACATACAGTCCCAAAAGCGCGG 90

seq_name: /SIDS5/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA148186

seq_documentation_block:

ID_AA148186 standard; DNA: 924 BP.

XX

AC_AA148186;

XX 17-OCT-2001 (first entry)

XX Probe #16872 used to measure gene expression in human placenta sample

XX Probe: microarray; human; placenta, antenatal diagnosis;
 genetic disorder; ss.

XX Homo sapiens.

XX W0200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001W0-NS00663

XX 04-FEB-2000; 2000NS-0180332.

XX 24-MAY-2000; 2000NS-0207456.

XX 30-JUN-2000; 2000NS-0608438.

XX 03-AUG-2000; 2000NS-0632356.

XX 21-SEP-2000; 2000NS-0234687.

XX 27-SEP-2000; 2000NS-0236359.

XX 04-OCT-2000; 2000NS-0024263

XX (MOLFE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for
 analyzing gene expression in human placenta -

claim 25; SEQ ID No 16872; 654pp; English.

XX the present invention relates to single exon nucleic acid probes (SPNs).

XX the present sequence is one such probe. The probes are useful for
 producing a microarray for predicting, measuring and displaying gene

XX expression in samples derived from human placenta. The probes are useful
 for antenatal diagnosis of human genetic disorders.

XX Sequence 924 BP; 163 A, 252 C, 310 G, 193 T, 0 other,

alignment_scores:

Quality: 155.00 Length: 27

Ratio: 5.741 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-557-262-4_copy_246_272 x AA148186/rev

Align seq 1/1 to reverse of: AA148186 from: 1 to: 924

1 PheYrPheAsnYsThrYsGlyPheCysLeuAraAspSerGlyAr 17
 |||||
 170 TTCTACTTTAAACAAACCAAGGGCTTTTACTGCTCGGACAGCGCGG 121

17 qAspAraCysLeuIleGlySerLysGlyArq 27
 |||||
 120 GACCGCGTGTACATACAGTCCCAAAAGCGCGG 90

seq_name: /SIDS5/gcdata/geneseq/geneseqn-emb1/NA1999.DAT:AAK47246

seq_documentation_block:

ID_AA37246 standard; DNA: 1305 BP.

XX

AC_AA37246;

XX 20-JUL-1999 (first entry)

XX Human 3-OST-1 encoding DNA.

XX

XX Heparan sulfate D-glucosaminyl 3-O-sulfotransferase; anticoagulant;
 saccharide; glycosaminoglycan; proteoglycan; antithrombin; 3-OST;
 heparan sulfate; thrombotic disorder; deep vein thrombosis;

XX pulmonary embolism; coagulation enzyme inactivation; ss.

XX Homo sapiens.

XX W09922005-A2.

XX 06-MAY-1999.

XX 23-OCT-1998; 98WO-US22597.

XX 31-OCT-1997; 97US-0065437.

XX 24-OCT-1997; 97US-0062762.

XX (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX

XX Fritze LMS, Liu J, Rosenberg RD, Schwartz JJ, Shworak NW;
 Zhang L;

XX WPI; 1999 312968/26.

XX FUSOH, AAV17063.

XX Heparan sulfate D-glucosaminyl 3-O-sulfotransferase and related
 polynucleotides

XX Claim 5; Page 74-75; 95pp; English.

XX the invention relates to nucleic acid molecules (AAK47245-X47250)
 encoding heparan sulfate D-glucosaminyl 3-O-sulfotransferases (3-OSTs).
 The 3-OST proteins can be used for 3-O-sulfating saccharide residues
 within a preparation of glycosaminoglycan or proteoglycan
 polysaccharides. 3-OST-1 can be used for enriching the
 antithrombin-binding fraction in a preparation of heparan sulfates (HS).
 3-OST-1 can also convert HS proteoglycan anticoagulant precursor to the
 HS anticoagulant (HSact). The 3-OSTs (optionally lacking enzymatic
 function) can be used to determine partial sequence information for
 complex polysaccharides. The 3-OST proteins, genes and antibodies are
 also useful for diagnosis of disorders involving HS biosynthesis. 3-OSTs
 are especially used to generate anticoagulant pentasaccharides, which may
 be used to treat thrombotic disorders such as deep vein thrombosis and
 pulmonary embolism. Coagulation enzyme inactivation by antithrombin is
 enhanced by complexing of antithrombin with endothelial cell surface HS
 proteoglycans. This is responsible for the non-thrombotic properties
 of blood vessels. The present sequence represents a human 3-OST-1
 encoding DNA.

XX Sequence 1305 BP; 304 A; 380 C; 129 G; 292 T; 0 other;

alignment_scores:

Quality: 155.00 Length: 27

Ratio: 5.741 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

CC pulmonary embolism. Coagulation enzyme inactivation by dactinomycin is
 CC enhanced by complexing of antithrombin with endothelial cell surface HS
 CC proteoglycans. This is responsible for the non-thrombotic properties
 CC of blood vessels. The present sequence represents a mouse 3-OST-1
 XX encoding DNA.

CC Sequence 1686 BP; 422 A; 426 C; 392 G; 456 T; 0 other;

alignment_scores:
 Quality: 152.00 Length: 27
 Ratio: 5.630 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:

US-09-557-262-4_COPY_246_272 x AAX17245

Align seq 1/1 to: AAX17245 from: 1 to: 1685

1 PheTyrPheAsnLysTyrLysGlyPheCysLeuArgAspSerGlyAr 17
 |||||
 1070 TTCTACTTAAACAAACCAACGGCTTCTACTGCTCGCGACAGTGGCAA 1119

17 gAspArgCysLeuHisGlySerLysGlyArq 27

|||||

1120 GAAAGAGTGTACAGAGTCAAGAGGCGGG 1150

seq_name: /SID55/gcdata/g/ncsseq/9-00-seq-emb1/NA2601B.DAT: ABL26623

seq_documentation_block:

ID ABL26623 standard; DNA; 897 BP.

XX ABL26623;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 31342.

XX Drosophila; developmental biology; cell signalling, insecticide;
 KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO US09231

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX Claim 1; SEQ ID NO 31342; 21pp + Sequence Listing, English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB16175), expressed DNA
 CC sequences (AB161840-AB16175) and the encoded proteins
 CC (AB57737-AB572072).

XX The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pcl_sequences.

XX Sequence 897 BP; 183 A; 233 C; 295 G; 186 T; 0 other;

alignment_scores:
 Quality: 114.00 Length: 27
 Ratio: 4.560 Gaps: 0
 Percent Similarity: 92.593 Percent Identity: 74.074

alignment_block:

us-09-557-262-4_COPY_246_272 x ABL26623

Align seq 1/1 to: ABL26623 from: 1 to: 897

1 PheTyrPheAsnLysTyrLysGlyPheCysLeuArgAspSerGlyAr 17
 |||||

703 TTTTACTTCAATGAGAGCAAGGGCTTCTACTGCTCGCGTATGCAACAGG 752

17 gAspArgCysLeuHisGlySerLysGlyArq 27

|||||

753 GCAATGCTGTCTCTGGACAGCAAGGGCGAGG 783

seq_name: /SID55/gcdata/g/ncsseq/geneseq/emb1/NA2601B.DAT: ABL26622

seq_documentation_block:

ID ABL26622 standard; DNA; 3484 BP.

XX ABL26622;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 31339.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX Claim 1; SEQ ID NO 31339; 21pp + Sequence Listing, English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB16175), expressed DNA
 CC sequences (AB161840-AB16175) and the encoded proteins
 CC (AB57737-AB572072).

XX The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pcl_sequences.

XX

CC complex polysaccharides. The 3-OST proteins, genes and antibodies are
 CC also useful for diagnosis of disorders involving HS biosynthesis. 3-OSTs
 CC are especially used to generate anticoagulant pentasaccharides, which may
 CC be used to treat thrombotic disorders such as deep vein thrombosis and
 CC pulmonary embolism. Coagulation enzyme inactivation by antithrombin is
 CC enhanced by complexing of antithrombin with endothelial cell surface HS
 CC proteoglycans. This is responsible for the non-thrombotic properties
 CC of blood vessels. The present sequence represents a human 3-OST-4
 CC encoding DNA.

SQ Sequence 3658 BP; 771 A, 1085 C; 1107 G; 695 T; 0 other;

alignment_scores:

Quality: 97.50 Length: 30
 Ratio: 4.239 Gaps: 1
 Percent Similarity: 76.667 Percent Identity: 66.667

alignment_block:

us-09-557-262-4_copy_246_272 x AAX37250 ..

Align seg 1/1 to: AAX37250 from: 1 to: 3658

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArg..... As 14
 |||||
 2002 TTCTATTTCACAAAGCAAGAGGCTTGGCTTAAGAGAGGCGAGAGA _2051
 14 pSerGlyArgAspArgCysGluHisGluSerLysGlyArg 27
 |||||
 2052 CAGCAGTCCCGGAGTCTTAGCGAGGAGGAGGAGGTCGG 2091

seq_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT: AAX37249

seq_documentation_block:

ID AAX37249 standard; DNA: 2032 BP.

XX AAX37249;

DT 20 JUL-1999 (first entry)

DE Human 3-OST-3B encoding DNA.

XX Heparan sulfate D-glucosaminyl 3-O-sulfotransferase; anticoagulant;
 KW saccharide; glycosaminoglycan; proteoglycan; antithrombin; 3-OST;
 KW heparan sulfate; thrombotic disorder; deep vein thrombosis;
 KW pulmonary embolism; coagulation enzyme inactivation; ss.

OS Homo sapiens.

PN W09922005-A2

XX 06-MAY-1999.

XX 23-OCT-1998; 98W0-0522597.

XX 31-OCT-1997; 97US-0065437.

XX 24-OCT-1997; 97US 0062762.

XX (MASI) MASSACHUSETTS INST TECHNOLOGY.

PI Fritze LMS, Liu J, Rosenberg PD, Schwartz JJ, Shworak NW;

PI Zhang L;

XX WPI; 1999-312968/26.

XX P-PSDB; AAY17066.

XX Heparan sulfate D-glucosaminyl 3-O-sulfotransferase and related

PT polynucleotides

XX Claim 5; Page 83-85; 95pp; English.

CC The invention relates to nucleic acid molecules (AAX37245-X37250)
 CC encoding heparan sulfate D-glucosaminyl 3-O-sulfotransferases (3-OSTs).

CC The 3-OST proteins can be used for 3-O-sulfating saccharide residues
 CC within a preparation of glycosaminoglycan or proteoglycan
 CC polysaccharides. 3-OST 1 can be used for enriching the
 CC antithrombin-binding fraction in a preparation of heparan sulfates (HS).
 CC 3-OST-1 can also convert HS proteoglycan anticoagulant precursor to the
 CC HS anticoagulant (HSa). The 3-OSTs (optionally lacking enzymatic
 CC function) can be used to determine partial sequence information for
 CC complex polysaccharides. The 3-OST proteins, genes and antibodies are
 CC also useful for diagnosis of disorders involving HS biosynthesis. 3-OSTs
 CC are especially used to generate anticoagulant pentasaccharides, which may
 CC be used to treat thrombotic disorders such as deep vein thrombosis and
 CC pulmonary embolism. Coagulation enzyme inactivation by antithrombin is
 CC enhanced by complexing of antithrombin with endothelial cell surface HS
 CC proteoglycans. This is responsible for the non-thrombotic properties
 CC of blood vessels. The present sequence represents a human 3-OST-3B
 CC encoding DNA.

SQ Sequence 2032 BP; 376 A; 673 C; 578 G; 405 T; 0 other;

alignment_scores:

Quality: 88.50 Length: 30
 Ratio: 3.848 Gaps: 1
 Percent Similarity: 76.667 Percent Identity: 60.000

alignment_block:

us-09-557-262-4_copy_246_272 x AAX37249 ..

Align seg 1/1 to: AAX37249 from: 1 to: 2032

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArg As 14
 |||||
 1306 TTCTATTTCACAAAGCAAGAGGCTTGGCTTAAGAGAGGCGAGGG 1355
 14 pSerGlyArgAspArgCysGluHisGluSerLysGlyArg 27
 |||||
 1356 CAGCAGTCCCGGAGTCTTAGCGAGGAGGAGGTCGG 1395

seq_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT: AAX37248

seq_documentation_block:

ID AAX37248 standard; DNA: 2314 BP.

XX AAX37248;

DT 20-JUL-1999 (first entry)

DE Human 3-OST-3A encoding DNA.

XX Heparan sulfate D-glucosaminyl 3-O-sulfotransferase; anticoagulant;
 KW saccharide; glycosaminoglycan; proteoglycan; antithrombin; 3-OST;
 KW heparan sulfate; thrombotic disorder; deep vein thrombosis;
 KW pulmonary embolism; coagulation enzyme inactivation; ss.

OS Homo sapiens.

PN W09922005 A2.

XX 06-MAY-1999.

XX 23-OCT-1998; 98W0-0522597.

XX 31-OCT-1997; 97US-0065437.

XX 24 OCT 1997; 97US 0062762.

XX (MASI) MASSACHUSETTS INST TECHNOLOGY.

PI Fritze LMS, Liu J, Rosenberg PD, Schwartz JJ, Shworak NW;

PI Zhang L;

XX WPI; 1999-312968/26

XX P-PSDB; AAY17065.

PT Heparan sulfate 6-glucosaminyl 3-O sulfotransferase and related
 PT polynucleotides
 XX
 PS Claim 5; Page 79-81; 95pp; English.
 XX
 CC The invention relates to nucleic acid molecules (AA37245-X37250)
 CC encoding heparan sulfate 6-glucosaminyl 3-O-sulfotransferases (3-OSTs).
 CC The 3-OST proteins can be used for 3-O sulfating saccharide residues
 CC within a preparation of glycosaminoglycan or proteoglycan
 CC polysaccharides. 3-OST-1 can be used for enriching the
 CC antithrombin-binding fraction in a preparation of heparan sulfates (HS).
 CC 3-OST-1 can also convert HS proteoglycan anticoagulant precursor to the
 CC HS anticoagulant (HSact). The 3-OSTs (optionally lacking enzymatic
 CC function) can be used to determine partial sequence information for
 CC complex polysaccharides. The 3-OST proteins, genes and antibodies are
 CC also useful for diagnosis of disorders involving HS biosynthesis. 3-OSTs
 CC are especially used to generate anticoagulant pentasaccharides, which may
 CC be used to treat thrombotic disorders such as deep vein thrombosis and
 CC pulmonary embolism. Coagulation enzyme inactivation by antithrombin is
 CC enhanced by complexing of antithrombin with endothelial cell surface HS
 CC proteoglycans. This is responsible for the non-thrombogenic properties
 CC of blood vessels. The present sequence represents a human 3-OST-3A
 CC encoding DNA.
 XX
 SQ Sequence 2314 BP: 442 A; 749 C; 742 G; 381 T; 0 other;

alignment_scores:

Quality:	88.50	Length:	30
Ratio:	1.848	Gaps:	1
Percent Similarity:	76.667	Percent Identity:	60.000

alignment_block:

US-09-557-262-4_COPY_246_272 x AA37248 ..
 Align seg 1/1 to: AA37248 from: 1 to: 2314
 1 PheTyrPheAsnTyrThrTyrGlyPheTyrCysLeuArg.....As 14
 |||||
 1819 TTCTACTTCACAAACACCAAGAGGCTTCCCTTCCTCAGCAAGAGGAGAGG 1868
 14 PSetClyArgAspArgCysLeuHISGluSerTyrGlyArg 27
 |||||
 1869 CAGCAGCGCGCGCCCACTGGCTTGGGCAAGACCAAGGCGCAGG 1908

OM of: US-09-557-262-4_copy_246_272.p2n.rge:Ent1 * out_format: fts

Date: Jul 17, 2002 8:58 AM

About: Results were produced by the GeneCore software, version 4.5.

Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet.p2n.model -DEV=xlp
-G=/cdm21/us09/spool/us0946/262-4_copy_246_272.p2n.rge -CAIOP=12.000
-DB=GenEmbl -QFMT=fastap -SUFFIX p2n.rge -GAIOP=12.000
-GAPEXT=4.000 -MINMATCH=6.100 -DEPEXT=6.000 -DEPEXT=6.000
-GAPEXT=4.500 -GAPEXT=6.050 -GAPEXT=10.000 -XGAPEXT=0.500
-DEPEXT=6.000 -DEPEXT=7.000 -GAPEXT=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DEPEXT=7.000 -START=1 -MATRIX=blossum62
-TRANS=human40.csi -LIST=45 -LOCALIGN=200 -THR_SCORE=pr+
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=fts
-NORM=ex1 -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09557262-4_COPY_246_272_COPY_272 -ICFU=3 -LONSLO3
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPEXT -WAIT -THPEADS=1

Search information block.

Query: US-09-557-262-4_COPY_246_272

Query length: 27

Database: GenEmbl.*

Database sequences: 1797656

Database length: 187333701

Search time (sec): 7102.980000

score_list:

Sequence	Strd	Orig	%Score	IScore	Len	Documentation
gb_pr:AF019386	+	155.00	422.73	3.5e-15	1305	AF019386 Homo sapiens heparan
gb_pr:AC073591	+	155.00	375.88	8.5e-13	159838	AC073591 Homo sapiens MA
gb_pr:AC006230	+	155.00	379.48	9.0e-13	166908	AC006230 Homo sapiens chro
gb_to:AF177436	+	152.00	417.85	7.2e-15	916	AF177436 Paltus norvegicus a
gb_to:BC009133	+	152.00	411.96	1.4e-14	1657	BC009133 Mus musculus, heparan
gb_to:AF019385	+	152.00	411.81	1.4e-14	1685	AF019385 Mus musculus, heparan
gb_btq:AC084416_3	+	152.00	374.56	1.7e-12	110000	Continuation (4 of 6) of AC0
gb_btq:AC098713	+	152.00	368.38	3.7e-12	220128	AC098713 Mus musculus chro
gb_in:AC004657	+	114.00	268.31	1.4e-06	76618	AC004657 Drosophila melanoga
gb_btq:AC018336	+	114.00	265.95	2.4e-06	110482	AC018336 Drosophila melanoga
gb_in:AC009328	+	114.00	263.31	4.0e-06	242458	AC009328 Drosophila melanoga
gb_in:AC003728	+	114.00	257.38	5.7e-06	261090	AC003728 Drosophila melanoga
gb_pr:AF019378	+	97.50	252.37	1.1e-05	2211	AF019378 Homo sapiens heparan
gb_pr:AC093511	+	97.50	212.72	0.0017	189234	AC093511 Homo sapiens chro
gb_btq:AC093359	+	97.50	211.43	0.0021	218502	AC093359 Mus musculus chro
gb_btq:AC021445	+	96.50	209.18	0.0027	203621	AC021445 Mus musculus clone
gb_pr:AF105377	+	88.50	227.20	0.0003	2532	AF105377 Homo sapiens heparan
gb_pr:AF105376	+	88.50	225.19	0.0004	2546	AF105376 Homo sapiens heparan
gb_pr:AK023723	+	88.50	224.31	0.0004	2408	AK023723 Homo sapiens cDNA Fl
gb_pr:AC005375	+	88.50	188.42	0.0369	149030	AC005375 Homo sapiens chro
gb_pr:AC005224	+	88.50	187.92	0.0419	166687	AC005224 Homo sapiens chro
gb_to:AF168992	+	87.50	226.09	0.0003	1656	AF168992 Mus musculus D-glyco
gb_pr:U54396	+	87.50	197.48	0.0124	41355	AC011723 Human DNA sequence
gb_pr:AC005163	+	87.50	192.14	0.0244	75108	AC005163 Homo sapiens chro
gb_pr:AC006640	+	87.50	191.40	0.0268	81574	AC006640 Homo sapiens chro
gb_btq:AC012180	+	87.50	181.35	0.0974	231997	AC012180 Homo sapiens chro
gb_btq:AL604889	+	87.50	179.22	0.1279	320096	AL604889 Mus musculus chro
gb_pr:AF107374	+	86.50	221.72	0.0005	1968	AF107374 Homo sapiens heparan
gb_pr:AF105375	+	86.50	221.72	0.0005	1968	AF105375 Homo sapiens heparan
gb_btq:AC027102	+	86.50	182.95	0.0893	168612	AC027102 Homo sapiens chro
gb_pr:HUAC02287	+	86.50	181.05	0.1011	188636	AC02287 Homo sapiens chro
gb_btq:AC098715	+	83.50	169.93	0.4213	249269	AC098715 Mus musculus chro
gb_btq:AX317978	+	75.50	189.39	0.0347	2115	AX317978 Human DNA sequen
gb_pr:AL355498	+	75.50	150.53	5.07	96745	AL355498 Human 1R sequence
gb_btq:AC017305	+	71.50	143.52	12.46	93745	AC017305 Drosophila melanoga
gb_btq:AC068459	+	71.50	139.27	21.49	93745	AC068459 Homo sapiens clone
gb_btq:AC108928	+	71.50	139.27	21.49	93745	AC108928 Homo sapiens clone
gb_btq:AC073025	+	71.50	138.86	22.65	948247	AC073025 Homo sapiens chro
gb_in:AC008337	+	71.50	138.78	22.87	169680	AC008337 Drosophila melanoga
gb_btq:AC005289	+	71.50	138.00	25.29	185314	AC005289 Paltus norvegicus c

gb_pr:AC006230 + 71.50 137.59 25.34 155.00 + AC006230 Homo sapiens chr
gb_in:AE003511 + 71.50 132.49 51.27 343807 + AE003511 Drosophila melan
gb_in:AB037943 + 69.00 179.05 0.4246 22565 + AB037943 Caenorhabditis ele
gb_in:AB037942 + 69.00 169.12 0.4673 2515 + AB037942 Caenorhabditis ele
gb_in:AB047441 + 69.00 168.67 0.4948 2644 + AB047441 Caenorhabditis ele

seq_name: gb_pr:AF019386

seq_documentation_block:
name: AF019386 type: mRNA size: 1305 bp map: chr14 chr14 NOV 1997
DEFINITION Homo sapiens heparan sulfate 3-O-sulfotransferase 1 precursor
ACCESSION AF019386
VERSION AF019386.1 GI:2618972
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1305)
AUTHORS Shworak, N.W., Liu, J., Fritz, L.M.S., Schwartz, J.J., Zhang, L.,
Logan, D., and Rosenberg, R.D.
TITLE Molecular cloning and expression of mouse and human cDNAs encoding
heparan sulfate D-glucosaminyl 3-O-sulfotransferase
JOURNAL J. Biol. Chem. 272 (44): 28068-28074 (1997)
MEDLINE 98010647
REFERENCE 2 (bases 1 to 1305)
AUTHORS Shworak, N.W., Liu, J., Fritz, L.M.S., Schwartz, J.J., Zhang, L.,
Logan, D., and Rosenberg, R.D.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-1997) Biology, MIT, 31 Ames Street, Cambridge, MA
02139, USA
FEATURES
source 1..1305
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Isolated from lambda triplex brain cDNA library
from Clontech"
gene 1..1305
/gene="3OST1"
5'UTR 1..118
/gene="3OST1"
3'-UTR 119..1305
/gene="3OST1"
CDS 119..1305
/gene="3OST1"
/function="rate limiting enzyme for synthesis of
anticoagulant heparan"
/note="heparan sulfate sulfotransferase; 3-OST1;
interluminal Gal41 resident protein (retension mechanism
unknown)"
/codon_start=1
/product="heparan sulfate 3 O sulfotransferase-1
precursor"
/protein_id="AA084388.1"
/db_xref="GI:2618973"
/translation="MAALLISAVILVAGQELVPSQAPLGGPFLPKAGTILQVPPVQ
VAPNSAQI POT IIGVKGGTHAIIEMI SIHPVAAAEVHFFDWERHYSGLQW
YLSQMFSPHGLTVKTPAYETSPKPYRVSMNPSTRLILIRDSRVISDYTOV
PYNMQHKFKFYSIEELVLRGFIKVYKALNPSLYHVMGNLFFELPHIVDHD
VLEFFETKVEFFLFTSTINASTYNTFTFTYTLFSSGSEFLHESKSHARPQ
VQFKLNLKHEVTEHFNKFFELVGTEDMI"
misc_feature 119..1309
/gene="3OST1"
misc_signal 260..268
/product="heparan sulfate 3 O sulfotransferase-1"
misc_feature 278..1039
/gene="3OST1"
/note="encodes potential N-linked site"
misc_signal 527..535
/note="encodes presumptive sulfotransferase catalytic
domain"


```

JOURNAL: Submitted (02 JUN 1999) Department of Genetics, Stanford Human
Genome Center, 4005 Miranda Avenue, Palo Alto, CA 94304, USA
COMMENT: On Jun 2, 1999 this sequence version replaced gi:4803914.
Quality: Phrap Quality >40 99.5% of Sequence;
Estimated Total Number of Errors is 0.3.
SIS Content:
SHGC-51595 G34132
WI-1308 G02734
SHGC-53120 G36506
SHGC-13860 G14876.
FEATURES             Location/Qualifiers
     source           1..166998
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4p16"
                     /clone="C0287314"
BASE COUNT          49526 a 33832 c 34587 g 49053 t
ORIGIN
1 PheTyRheAsnLysThrLysGlyPheTyRheCysLeuArgAspSerGlyAr 17
101629 TTTCTACITTTACAAACCAAGCGGCTTTATCGTGTGGGACACCGCGG 101580
17 gAspArgCysLeuHisGluSerLysGlyArq 27
101579 GGACGGCTGTTCATGATGATGCAAGGCGCG 101549
seq_name: qb_ro:AF177430
seq_documentation_block:
LOCUS       AF177430                936 bp  mRNA  linear  pop 10-MAY-2001
DEFINITION  Rattus norvegicus 3-O-sulfotransferase mRNA, complete cds.
ACCESSION   AF177430
VERSION     AF177430.1 GI:9957243
KEYWORDS    Norway rat.
SOURCE      Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 936)
AUTHORS    Li,Z.Y., Hirayoshi,K. and Suzuki,Y.
TITLE      Expression of N-deacetylase/sulfotransferase and
            3-O-sulfotransferase in rat alveolar type II cells
JOURNAL    Am. J. Physiol. 279 (2), L592-L595 (2000)
REFERENCE   2 (bases 1 to 936)
AUTHORS    Li,Z.Y., Hirayoshi,K. and Suzuki,Y.
TITLE      Direct Submission
JOURNAL
SUBMITTED  (13-AUG-1999) Department of Ultrastructural Research,
            Institute for Frontier Medical Sciences, 53 Shogoin Kawahara-cho,
            Sakyo-ku, Kyoto 606-8507, Japan
FEATURES             Location/Qualifiers
     source           1..936
                     /organism="Rattus norvegicus"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
                     /cell_type="type II epithelium"
                     1..936
                     /note="3-OST"

```

```

/codon_start=1
/product="3-O-sulfotransferase"
/protein_id="AAC09284.1"
/db_xref="GI:9957244"
/translation="MTITLIGAVLIVAGPQIVPSIPAAPDPRIKQQLIPKVIILPEL
TGGGAATNGSTQQLPQTIIIGVPKGGTALLLEMLSTHDDVAANENVFHFWERHSQ
GLGWYLTQMPFSSPHQIVKIPAYETSPKVERIHSMNPTIRLLILLRDSKPLVLSQ
YTQVLYNHLQKHKPYPPFDLLMRUCRINVOYKALNLSLYHAHMLNMLRPPDGLGHIHI
VDRRTFRTPPEIKGVFFLLTSPETNASNFYKRTGFGYCLDSGSDKCLUESKGR
AHPQVDPKLLDKLHEYFREPNNKFKFKLGRTEEDWI"
BASE COUNT          224 a 292 c 224 g 196 t
ORIGIN
1 PheTyRheAsnLysThrLysGlyPheTyRheCysLeuArgAspSerGlyAr 17
748 TTTCTATTTTAAATAAAATAAGGATPATAATGCTGTGGGACASTGGCA 797
17 gAspArgCysLeuHisGluSerLysGlyArq 27
798 GCACCGCTGTTCATGATGATGCAAGGCGCG 828
seq_name: qb_ro:BC009133
seq_documentation_block:
LOCUS       BC009133                1657 bp  mRNA  linear  REF 12-JUN-2001
DEFINITION  Mus musculus, heparan sulfate (glucosamine) 3-O-sulfotransferase 1,
            clone MGC:11450 (IMAGE:3155049), mRNA, complete cds.
ACCESSION   BC009133
VERSION     BC009133.1 GI:1414662
KEYWORDS    house mouse.
SOURCE      Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
            Strausberg,R.
REFERENCE   1 (bases 1 to 1657)
AUTHORS    Strausberg,R.
TITLE      Direct Submission
JOURNAL
SUBMITTED  (05-JUN-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892 2590,
            USA
            NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk
            Email: mgc@bbs-research.nih.gov
            Tissue procurement: Iohar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgc.bcm.tmc.edu/cdna/
            Contact: villalona@bcm.tmc.edu.
            Villalón, D.K., Luna, P.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
            A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
            Muzny,D.M., Gibbs,R.A.
Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Classification at http://image.llnl.gov
            Series: IRAP Plate: 16 Row: O Column: 17
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA (gi: 2618970).
            Location/Qualifiers

```


source	1. .1657
	/organism="Mus musculus"
	/db_xref="locus:15476"
	/db_xref="taxon:10090"
	/clone "MCC:11450 IMAGR:3155049"
	/tissue_type="mammary tumor, Brca1+/fl; MMTV-Cre model, 10 months old, gross tissue"
	/clone_lib="NCL-CGAP_Mam3"
	/lab_host="DH10B"
	/note="Vector: pCMV-SPORT6"
	284 1219
CDS	/codon_start=1
	/product="heparan sulfate (glucosamine)
	3-O-sulfotransferase 1"
	/protein_id="AA09133.1"
	/db_xref="GI:14318663"
	/translation="MTLLLGAVLLVAQPOLVHSPAAPGPGGLKQELLRKVILHPED
	TGEGTASNGTQGLPVTIIIGVRKGTTRALLEMLSLHPDVAANAENEVHFFWEERYSQ
	GLWYITQMPFSSPHQIIVHKITAYFTSPKVPRIHSMNPTIRLLILRDPSEVLSQ
	VTQVLNHLQKHKPYPIEDLLMRDGLNLDYKALNRSLYHAHMLNLRFFPLGHIHI
	VDGDIILRDPEPELOKVPFKILSPQINASNYFNKTKGYCLPISQKPCGLHFSKGR
	AHPQVDPKLLDKLHEYFHEPNKFEKLVGRTDWH"
BASE COUNT	398 a 439 c 393 g 427 t
ORIGIN	AHPQVDPKLLDKLHEYFHEPNKFEKLVGRTDWH"
alignment_scores	Quality: 152.00 length: 27
	Ratio: 5.630 Gaps: 0
Percent Similarity:	100.000 Percent Identity: 96.296
alignment_block	US-09-457-252-4_2013_246_272 x NC009133 ..
Align seq 1/1 to NC009133 from 1 to 1657	
1	PhETyPheASLysThrLysGlyPheTyrcysLeuArqAspSerGlyAr 17
1031	TTCACCTTTTAAACAAACCAAGGCGTCTACTGCGTGGGACAGTGGCAA 1080
17	gaspArGcysLeuHisGluSerLysClyArg 27
1081	GAGCGCTGCTTAAAGATGTAAGAAAGACGG 1111
seq_name: gb_ro:AF019385	
seq_documentation_block:	
LOCUS	AF019385 1685 bp mRNA linear 100 14-NOV-1997
DEFINITION	Mus musculus heparan sulfate D-glucosaminyl 3-O-sulfotransferase-1 precursor (30ST1) mRNA, alternatively spliced, complete cds.
ACCESSION	AF019385
VERSION	AF019385.1 GI:2618970
KEYWORDS	
SOURCE	house mouse.
ORGANISM	Mus musculus
	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE	1 (bases 1 to 1685)
AUTHORS	Shworak, N. W., Liu, J., Fritze, L. M. S., Schwartz, T. T., Zhang, L.,
	Loefer, D., and Rosenberg, R. D.
TITLE	Molecular cloning and expression of mouse and human cDNAs encoding
	heparan sulfate D-glucosaminyl 3-O-sulfotransferase
JOURNAL	J. Biol. Chem. 272 (44), 28068-28079 (1997)
MEDLINE	98010647
REFERENCE	2 (bases 1 to 1685)
AUTHORS	Shworak, N. W., Liu, J., Fritze, L. M. S., Schwartz, T. T., Zhang, L.,
	Loefer, D., and Rosenberg, R. D.
TITLE	Direct Submission
JOURNAL	Submitted (14 AUG 1997) Biology, MIT, 31 Ames Street, Cambridge, MA
FEATURES	02139, USA
	location/Qualifiers
source	1. .1657
	/organism="Mus musculus"
	/strain="C3H/An"
	/db_xref="taxon:10090"
	/cell_line="clonal 1 cell line LTA clone 1A (described in
	Shworak et al., 1996, J. Biol. Chem. 271-27071-82)"
	/note="poly(A)+ KNA isolated from postconfluent cultures"
5'UTR	1. .322
	/gene="30st1"
	/note="alternatively spliced; this version represents a
	Class 1 mRNA"
gene	1. .1685
	/gene="30st1"
	216. .217
misc_feature	/gene="30st1"
	/note="splice variant site; many additional 3'OST-1 cDNAs
	(Class 2 to Class 9) contain extra 5' UTR exons between
	these nucleotides (extra 86 to 1629 bp)"
sig_peptide	323. .382
	/gene="30st1"
CDS	323. .1258
	/gene="30st1"
	/function="rate limiting enzyme for the synthesis of
	anticoagulant heparan sulfate"
	/note="heparan sulfate sulfotransferase; 3-OST-1;
	interluminal Golgi resident protein (retension mechanism
	unknown)"
	/codon_start=1
	/product="heparan sulfate D-glucosaminyl
	3-O-sulfotransferase-1 precursor"
	/protein_id="AA84387.1"
	/db_xref="GI:2618971"
	/translation="MTLLLGAVLLVAQPOLVHSPAAPGPGGLKQELLRKVILHPED
	IGEGTASNGTQGLPVTIIIGVRKGTTRALLEMLSLHPDVAANAENEVHFFWEERYSQ
	GLWYITQMPFSSPHQIIVHKITAYFTSPKVPRIHSMNPTIRLLILRDPSEVLSQ
	VTQVLNHLQKHKPYPIEDLLMRDGLNLDYKALNRSLYHAHMLNLRFFPLGHIHI
	VDGDIILRDPEPELOKVPFKILSPQINASNYFNKTKGYCLPISQKPCGLHFSKGR
	AHPQVDPKLLDKLHEYFHEPNKFEKLVGRTDWH"
	383. .1255
mal_peptide	/gene="30st1"
	/note="confirmed by N-terminal analysis"
	/evidence=experimental
	/product="heparan sulfate D-glucosaminyl
	3-O-sulfotransferase-1"
misc_signal	476 484
	/gene="30st1"
	/note="encodes N-linked glycosylation sites"
	/evidence=experimental
misc_feature	494. .1255
	/gene="30st1"
	/note="encodes presumptive sulfotransferase catalytic
	domain"
misc_signal	908. .916
	/gene="30st1"
	/note="encodes N-linked glycosylation sites"
	/evidence=experimental
misc_signal	1058. .1066
	/gene="30st1"
	/note="encodes N-linked glycosylation sites"
	/evidence=experimental
misc_signal	1079. .1087
	/gene="30st1"
	/note="encodes N-linked glycosylation sites"
	/evidence=experimental
misc_structure	1100 1129
	/gene="30st1"
	/note="encodes cysteine bridged peptide loop"
	/evidence=experimental
3'UTR	1259. .1285
	/gene="30st1"
	1532. .1537
polyA_signal	/gene="30st1"
	1556
polyA_site	/gene="30st1"
polyA_signal	1586. .1591

polya_signal /gene-"30st1"
1530..1595
polya_site /gene-"30st1"
1605
polya_site /gene-"30st1"
1610
polya_signal /gene-"30st1"
1611..1615
polya_site /gene-"30st1"
1629
polya_signal /gene-"30st1"
1667..1672
polya_site /gene-"30st1"
1685
polya_site /gene-"30st1"
BASE COUNT 422 a 426 c 382 g 455 t
ORIGIN

alignment_scores:
Quality: 152.00 Length: 27
Ratio: 5.630 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:
US-09-557-262-4_copy_246_272 x AF019385
Align seq 1/1 to: AF019385 from: 1 to: 1685

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyAr 17
1070 TTTCTACTTTAAACAAACCAACAGCTTCTACTGCGTGGCGACAGTGGCAA 1119
17 gAspArgCysLeuHisGluSerLysGlyArg 27
1120 GGACCGCTTGCCTTACACAGTCAAAAGCCCGG 1150
seq_name: qb_btq.AC084416_3

seq_documentation_block:
WPCOMMENT
Sequence split into 6 fragments LOCUS AC084416 Accession AC084416

Fragment Name	Begin	End
AC084416_0	1	110000
AC084416_1	100001	210000
AC084416_2	200001	310000
AC084416_3	300001	410000
AC084416_4	400001	510000
AC084416_5	500001	595289

Continuation (4 of 6) of AC084416 from base 300001 (AC084416 from base 300001 to 1100000

alignment_scores:
Quality: 152.00 Length: 27
Ratio: 5.630 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:
US-09-557-262-4_copy_246_272 x AC084416_3/rev
Align seq 1/1 to reverse of: AC084416_3 from: 1 to: 1100000
1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyAr 17
7710 TTTCTACTTTAAACAAACCAACAGCTTCTACTGCGTGGCGACAGTGGCAA 77081
17 gAspArgCysLeuHisGluSerLysGlyArg 27
77080 GAAAGCTGATTAAGTCAAAAGCCCGG 77050
seq_name: qb_btq.AC098713

seq_documentation_block:
WPCOMMENT

LOCUS
DEFINITION
AC098713
AC098713
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC098713 220128 bp DNA Project HPC 27 Nov 2001
Mus musculus chromosome UNK clone RP23-2C16, WORKING DRAFT
SEQUENCE, 6 unordered pieces.
AC098713
AC098713 GI:17105314
HUG: HUGS_PHASE1; HUGS_DRAFT; HUGS_ACTIVEFIN.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 220128)
McPherson, J.D. and Waterston, R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 220128)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (31-OCT-2001) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Nov 27, 2001 this sequence version replaced 4117017621.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/seq/index.shtml
Contact: submissions@watson.wustl.edu
----- Project Information
Center project name: M.BA0002C16

----- Summary Statistics
Sequencing vector: M13; 27x
Sequencing vector: plasmid; 73x
Chemistry: Dye-primer ET; 0% of reads
Assembly: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.940319
Consensus quality: 217754 bases at least Q40
Consensus quality: 218359 bases at least Q30
Consensus quality: 218787 bases at least Q20
Insert size: 1; agarose-1p
Insert size: 221598; sum-of-contin
Quality coverage: 26.01 in Q20 bases; agarose 1p
Quality coverage: 23.96 in Q20 bases; sum of contin

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1043: contig of 1043 bp in length
* 1044 1143: gap of unknown length
* 1144 4765: contig of 3622 bp in length
* 4766 4865: gap of unknown length
* 4866 38329: contig of 33464 bp in length
* 38330 38429: gap of unknown length
* 38430 218537: contig of 180108 bp in length
* 218538 218637: gap of unknown length
* 218638 219339: contig of 702 bp in length
* 219340 219439: gap of unknown length
* 219440 220128: contig of 689 bp in length.

FEATURES
source
1..220128
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-2C16"
1..1043
/note="assembly_name:Contig16"
1144..4765
/note="assembly_name:Contig86"
4866..38329
misc_feature
1..1043
misc_feature
1144..4765
misc_feature
4866..38329

misc_feature /note="assembly_name:Contig87"
38430..218537
misc_feature /note="assembly_name:Contig88"
218638..219339
misc_feature /note="assembly_name:Contig43"
219440..220128
BASE COUNT 64542 a 44187 c 44623 g 66244 t 527 others
ORIGIN

alignment_scores:
Quality: 152.00 Length: 27
Ratio: 5.630 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:

US-09-557-262-4_copy_246_272 x AC004657 112668

Align seg 1/1 to: AC004657 from: 1 to: 220128

1 PhcAAsnLysThrLysGlyPheCysLeuArgAspSerGlyAr 17
TTTTTATTTTAA/AAAA/AAAGGATTTCTACTATGCTATGAGTAA 112668

17 gAspArqCysLeuHisGluSerLysGlyArq 27

CGACCGCTCTCTACACGACGCTCCAAAGCGCGC 112668

seq_name: gb_in:AC004657

seq_documentation_block:

LOCUS AC004657 76618 bp DNA linear INV 02 MAY-1998
DEFINITION Drosophila melanogaster DNA sequence (p1s DS08204 (D183) and DS09119 (D236)), complete sequence.

ACCESSION A-004657 A-004297 A-004347 A-004377 A-004382 A-004384 A-004378
VERSION AC003479 AC004338

KEYWORDS AC004657.1 GI:3108014

SOURCE HTS.

ORGANISM Drosophila melanogaster (Subclones in Sac from p1 clones DS08204 (D183) and DS09119 (D236)) DNA.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyceta; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 76618)

Celniker, S. F., Aghavani, A., Arcaina, T. T., Baxter, F., Blazek, P. G., Chew, M., Doyle, C. M., Parfan, D. E., Planagan, J., Houston, K. A., Hummasti, S. R., Karra, K., Kearney, L., Kim, S. H., Lee, B., Lomotan, M. A., Mazda, P., Mok, M. S., Moshrefi, A. R., Moshrefi, M., Nixon, K., Pacleb, J. M., Park, S., Pfeiffer, B., Pouch, D., Snir, E., Twomey, B., Wan, K. H., Whitehead, K. R., Yee, A., Zhang, R., Zieran, L. L., and Kimmel, B.

Sequencing of Drosophila chromosome 2R, region 55E9-55F2

TITLE

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 76618)

Celniker, S. F., Aghavani, A., Arcaina, T. T., Baxter, E., Blazek, P. G., Chew, M., Doyle, C. M., Parfan, D. E., Planagan, J., Houston, K. A., Hummasti, S. R., Karra, K., Kearney, L., Kim, S. H., Lee, B., Lomotan, M. A., Mazda, P., Mok, M. S., Moshrefi, A. R., Moshrefi, M., Nixon, K., Pacleb, J. M., Park, S., Pfeiffer, B., Pouch, D., Snir, E., Twomey, B., Wan, K. H., Whitehead, K. R., Yee, A., Zhang, R., Zieran, L. L., and Kimmel, B.

Direct Submission

TITLE

JOURNAL

REFERENCE

AUTHORS

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

For further information about this sequence, including its location

and relationship to other sequences, please visit our sequence

archive Web site
(http://fruitfly.berkeley.edu/sequence-archive.html) or send email
to drosophila@hgsc.lbl.gov.

Library Locations: 44, 86, 138, 95.

FEATURES

source

1..76618 Drosophila melanogaster

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/chromosome="2R"

/map="55E9-55F2"

/clone="p1s DS08204 (D183) and DS09119 (D236)"

/note="D236 extends from bp 1 to bp 35067 and D183 extends

from bp 35068 to bp 76618."

RASF COUNT 21623 a 16343 c 16625 g 22029 t

ORIGIN

alignment_scores:

Quality: 114.00 Length: 27

Ratio: 4.560 Gaps: 0

Percent Similarity: 92.593 Percent Identity: 74.074

alignment_block:

US-09-557-262-4_copy_246_272 x AC004657

Align seg 1/1 to: AC004657 from: 1 to: 76618

1 PhcAAsnLysThrLysGlyPheCysLeuArgAspSerGlyAr 17

TTTTTATTTTAA/AAAA/AAAGGATTTCTACTATGCTATGAGTAA 71708

71708 TTTTATTTTAA/AAAA/AAAGGATTTCTACTATGCTATGAGTAA 71757

17 gAspArqCysLeuHisGluSerLysGlyArq 27

CGACCGCTCTCTACACGACGCTCCAAAGCGCGC 71758

seq_name: gb_in:AC004657

seq_documentation_block:

LOCUS AC004657 76618 bp DNA linear INV 09-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCE IN PROGRESS ***. In ordered pieces.

ACCESSION AC004657

VERSION AC004657.1 GI:5552865

KEYWORDS HTS; HTGS; PHASFP

SOURCE fruit fly

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachyceta; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 110482)

Adams, M. and Venter, J. C.

Direct Submission

Submitted (09-NOV-1999)

Rockville, MD, USA

This sequence was identified as CDM:10214359 by the submitter.

For more information on this record email to flybase-l@ra.com.

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

Location/Qualifiers

1..110482

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

BASE COUNT 30512 a 25052 c 24519 g 30499 t

ORIGIN

alignment_scores:

Quality: 114.00 Length: 27

Ratio: 4.560 Gaps: 0

Percent Similarity: 92.593 Percent Identity: 74.074

/translaton="MK;PNIKKPIAVVLTISFHTKPIIAFKPKGNAFIVKUSTINI
 I PNPFTIPNPFNFNFPPDYPYGHIESPMSCSPFGGGGAGYSNSGASNSRGVINPN
 YNSGNSGNSGNSGLDRYSYSGNSGNSGNGNPNFNSYASNSGNSGNSGNSG
 GNSGNSGNSGNSGNSGNSGNSGNSGNSGNSGNSGNSGNSGNSGNSGNSG
 GNSGNSGNSGNSGNSGNSGNSGNSGNSGNSGNSGNSGNSGNSGNSGNSG
 THGGRDPCHPHDKPLPSRGPSPSAGOSI.PKYGNNTAVGGGYLYOKKOOGCAPCGKA
 SI.SKNWSTISORHGOOGAPQHQCTIHWLGGGSGSOYADGGAAGAAGGGOYWAYNMGSY
 ENGAUGAUAUUGUUUMMSMMWCUCGUSGAAACNAAGSSSVNVGGGANNUGGAANH
 YWSJASTSNAGADRK"
 complement(join(6554..6694,6763..7244,7311..8162,
 8270..9040,9042..9069))
 /gene="CG5477"
 /product="CT1730"
 /db_xref="FLYBASE:Fban0005477"
 /db_xref="FLYBASE:Fban00034370"
 complement(-6554..9069)
 /gene="CG5477"
 /map="5E5-55B6"
 /db_xref="FLYBASE:Fban0005477"
 /db_xref="FLYBASE:Fban0034370"
 complement(join(6554..6694,6763..7244,7311..8162,
 8270..9040,9042..9069))
 /gene="CG5477"
 /note="CG5477 gene product"
 /codon_start=1
 /db_xref="FLYBASE:Fban0005477"
 /db_xref="FLYBASE:Fban0034370"
 /protein_id="AAF57680.1"
 /db_xref="GI:7302578"
 /translaton="MIVAKILPKMKVVDLPHLSQSLDLPKGVKXAVILERLAYVEGZA
 CGDNNAVTPSRKRTIKRSRSPVOAAFAARPAVDITLEEGDEPDKLVIPDPEE
 SEQPAAVFPQSFEPRFAAAAYETDIVNOANRESGEPDEFKSKTIDRKQRTI
 EEAVPVVGVEADEMEEDHDAAPQEQTOTEEPEEKPAESTVAESNDSDJK
 MDVEDSAAPKTAEETEPRAKPDEOPBPBKSIHSBSKSPSGSGSPKHRSAGVKG
 PESKAANYVPDEPTIEENKVYLWNLSLDLPIDPTTFASAPLTSEITSLSLWS
 GARANYVRGKVFVRLSHSPVENSHTLIHPFHVGFPGFVGFSPKSSITLGEAPH
 STYELCPKALQSFLLFYKPYLGGVWDQYYLTSFTPIINTINCEDIQVAFPE
 KSLIHGSAKPLHVILKPYLVNLSQLGLVNAIRPKPKPKKDKKDDKDIDKI
 NDNGEKMLDEATADDVEKDDKESKEDEEAEAPERAKEDAAAPTETEAE
 TKVAPAANAAPSEATEASTETSKSETAAVSNGDAAAQAQDEXKASEDUKPS
 ENNEDEDGPSNKRPTDGDSKASEKEKORSQTTTEDEVVVPFPPTAALLIG
 VVLGLVPPKLVSQPORVSKRCFVILLVGLPCACKTHAHKHVAFNAUKYELLIG
 PDALISKMTDCASRKLVIHGWRDVKYETHICINSIALLEDIAMRRKRNFIIDOVNH"
 complement(join(11870..11995,12058..12219,12282..12490,
 12555..12737,12828..13185,28502..328622))
 /gene="sp2637"
 /product="CT142573"
 /db_xref="FLYBASE:Fban0005473"
 /db_xref="FLYBASE:Fban0034371"
 complement(join(11870..11995,12058..12219,12282..12490,
 12555..12737,12828..13185,28546..323644))
 /gene="sp2637"
 /product="CT17346"
 /db_xref="FLYBASE:Fban0005473"
 /db_xref="FLYBASE:Fban0034371"
 complement(join(11870..11995,12058..12219,12282..12490,
 12555..12737,12828..13185,28546..323644))
 /gene="sp2637"
 /note="CG5473"
 /map="5E5-55B6"
 /db_xref="FLYBASE:Fban0005473"
 /db_xref="FLYBASE:Fban0034371"
 complement(join(11870..11995,12058..12219,12282..12490,
 12555..12737,12828..13185,28546..323644))
 /gene="sp2637"
 /note="sp2637 gene product [alt 2]"
 /codon_start=1
 /db_xref="FLYBASE:Fban0005473"
 /db_xref="FLYBASE:Fban0034371"
 /protein_id="AAF57658.1"
 /db_xref="GI:7302576"
 /translaton="MVIVINLVLTIDPMDINSIFLIDPVVYVAYALULSLIQASVSP
 VGLTVGGFRMAASAPKHKVNYTCAGATTCTIVVVGSGSGSAVALAFRDGSGVIRA
 VGTVMYRVQPLTAGVPGCFPIQLIGGPIAKGYGHDVFFISIMPSHNHLIELQLQA

Quality levels above 40 are expected to have less than 1 error in 10,000 bp
 Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero.

Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

FEATURES
 source 1..189234
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="CID-2507C6"
 BASE COUNT 55941 a 38428 c 38216 g 56658 t 1 others
 ORIGIN

alignment_scores:
 Quality: 97.50 Length: 30
 Ratio: 4.239 Gaps: 1
 Percent Similarity: 76.667 Percent Identity: 66.667

alignment_block:
 US-09-557-262-4_copy_246_272 x AC093351/rev
 Align seg 1/1 to reverse of: AC093351 from 1 to 189234

1 PheTyPheAsnLysThrLysClyPheLysCysLeuArg.....AS 14
 114117 TCTATTTCACAAACCAAGGAGGTTTCCTTAAAGAAAGAGAAGA 114068
 14 pSerGlyArgAspArgCysLeuHisGluSerLysGlyArg 27
 114067 CAGCAGTCCCGAGGCTCTTAGGCAAGCAAGAGGTCGG 114028

seq_name: qb_hfq.AC093359

seq_documentation_block.
 LOCUS AC093359 218502 bp RNA linear RFC 20-AUG-2001
 DEFINITION Mus musculus clone RP23-57A19, WORKING DRAFT SEQUENCE, 16 unordered
 pieres
 AC093359
 VERSION AC093359.1 GI:15213904
 KEYWORDS HTG: HTGS_PHASE1: HTGS_DPAPT: HTGS_PHLITAP
 SOURCE house mouse.
 ORGANISM Mus musculus
 Pukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi,
 Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
 1 (bases 1 to 218502)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Mus musculus, clone RP23-57A19
 Unpublished
 2 (bases 1 to 218502)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Bonislawski,T., Bonkhalter,R.,
 Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
 Choquet,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., DeAreliauo,K., Dewar,K., Diaz,J.S., Dady,S., Fato,S.,
 Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Glade,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Haqos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
 Jones,C., Kamal,A., Karalis,A., Kells,C., Larocque,K.,
 Lamazares,R., Landers,T., Lebecky,J., Levine,R., Liu,G.,
 Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,

TITLE
 JOURNAL
 COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIRB
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: 114290
 Center clone name: 57_A_19

Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: dye terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 21338 bases at least Q40
 Consensus quality: 21548 bases at least Q40
 Consensus quality: 21643 bases at least Q20
 Insert size: 235000, agarose-lp
 Insert size: 217002; sum of contigs
 Quality coverage: 8.7 in Q20 bases; 9.4 in Q20 bases; sum of contigs

NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 59151: contig of 59151 bp in length
 59152 59251: gap of 100 bp
 59252 6031: contig of 1860 bp in length
 6032 60411: gap of 100 bp
 60412 62063: contig of 1652 bp in length
 62064 62163: gap of 100 bp
 62164 64111: contig of 1948 bp in length
 64112 64211: gap of 100 bp
 64212 65793: contig of 1582 bp in length
 65794 65893: gap of 100 bp
 65894 69038: contig of 3145 bp in length
 69039 69138: gap of 100 bp
 69139 74310: contig of 5172 bp in length
 74311 74410: gap of 100 bp
 74411 82566: contig of 8156 bp in length
 82567 82666: gap of 100 bp
 82667 93394: contig of 10728 bp in length
 93395 93494: gap of 100 bp
 93495 105599: contig of 12105 bp in length
 105600 105699: gap of 100 bp
 105700 117559: contig of 11860 bp in length
 117560 117559: gap of 100 bp
 117660 337463: contig of 19804 bp in length
 137464 137563: gap of 100 bp
 137564 157070: contig of 19507 bp in length
 157071 157170: gap of 100 bp
 157171 177936: contig of 20766 bp in length
 177937 178036: gap of 100 bp

* 178037 205897: contig of 27861 bp in length
 * 205898 205997: gap of 100 bp
 * 205998 218502: contig of 12505 bp in length.

FEATURES
source

Location/Qualifiers
 1..218502
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="RP24-57A19"
 /clone_lib="RPCL-23 Female Mouse BAC"
 1..59151
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 59252..60311
 /note="assembly_fragment"
 60412..62063
 /note="assembly_fragment"
 62164..64111
 /note="assembly_fragment"
 64212..65793
 /note="assembly_fragment"
 65894..69048
 /note="assembly_fragment"
 69139..774310
 /note="assembly_fragment"
 74411..82566
 /note="assembly_fragment"
 82667..93394
 /note="assembly_fragment"
 93495..110559
 /note="assembly_fragment"
 105700..117559
 /note="assembly_fragment"
 117660..137464
 /note="assembly_fragment"
 137564..157070
 /note="assembly_fragment"
 157171..177936
 /note="assembly_fragment"
 178037..205897
 /note="assembly_fragment"
 205998..218502
 /note="assembly_fragment"
 clone_end:17
 vector_side:right"
 BASE COUNT 64182 a 43359 c 44675 g 64786 t 1500 others
 ORIGIN

alignment_scores:
 Quality: 97.50 Length: 30
 Ratio: 4.239 Gaps: 1
 Percent Similarity: 76.667 Percent Identity: 66.667

alignment_block:

US-09-557-262-4_COPY_246_272 x AC093359

Align seq 1/1 to: AC093359 from: 1 to: 218502

1 phtyrlpAsnTyrLysGlyPheTyrCysLeuArg.....As 14
 |||||
 35843 TTATTACTCAACAAACCAAGCGTTCCTCCCTCCCTCAAGAGCCACAGCA 35892
 |||||
 14 pSerGlyArgAspArgCysLeuHisGluSerLysGlyArg 27
 |||||
 35893 CAGCAGCGCCGCAAGATGTTGAGCAAGCAAGCAAGGTCGG 35932

Journal Pre-proof

alignment_block:

US-09-557-262_4_COPY_21_303 x AA337246

Align seq 1/1 to: AA337246 from: 1 to: 1305

```

1  ArgProAlaGluLeuGlyGlnGlnGlnLeuLeuArgGlySerAlaGlnGlnL 17
|||||
179 GCGCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 228
|||||
17  uGluAspAspValArgAspGlyValAlaProAsnGlySerAlaGlnGlnL 34
|||||
229 CCGACATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 278
|||||
34  euProGlnThrIleIleIleGlyValArgGlyGlyGlyThrArgAlaLeu 50
|||||
279 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 328
|||||
51  LeuGluMetLeuSerLeuHisProAspValAlaAlaAlaGluAsnGluVa 67
|||||
329 CCGACATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 378
|||||
67  IuIsPhePheAspTrpGluGlnHisTyrSerHisGlyLeuGlyTrpTrpL 84
|||||
379 CCACATCTTCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 428
|||||
84  euSerGlnMetProPheSerTrpProHisGlnLeuThrValGlnGlyThr 100
|||||
429 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 478
|||||
101 ProAlaTyrPheThrSerProIysValProGluArgValTyrSerMetAs 117
|||||
479 CCGCGCGTAATTCAGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 528
|||||
117 nProSerIleArgLeuLeuLeuLeuArgAspProSerGluArgValL 134
|||||
529 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 578
|||||
134 euSerAspTyrThrGlnValIlePheTyrAsnHisMetGlnIlyshisysPro 150
|||||
579 TATCTGATATATATATATATATATATATATATATATATATATATATAT 628
|||||
151 TyrProSerIleGluGluPheLeuValArgAspGlyValArgAsnValAs 167
|||||
629 TACCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 678
|||||
167 pTyrLysAlaLeuAsnArgSerLeuTyrHisValHisMetGlnAsnTrpC 184
|||||
679 CTACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 728
|||||
184 euArgPhePheProLeuArgHisIleHisIleValAspGlyAspArgLeu 200
|||||
729 TCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 778
|||||
201 IleArgAspTrpPheLeuGluGlnGlyValGluArgPheLeuLysLe 217
|||||
779 ATCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 828
|||||
217 uSerProGlnIleAsnAlaSerAsnThrTyrPheAsnLysThrLysGlyP 234
|||||
829 CTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 878
|||||
234 heTyrCysLeuArgAspSerGlyValArgAspArgCysLeuHisGluSer 250
|||||
879 TTTTATCTGATATATATATATATATATATATATATATATATATATATAT 928
|||||
251 GlyArgAlaHisProGlnValAspPheLysLeuLeuAsnLysLeuHisG 267
|||||
929 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 978
|||||
267 uTyrPheHisGlnProAsnArgValTyrPheThrGlnLeuValGlyArgThr 283
|||||
979 ATATTTTCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1027

```

seq_name: /SID5/4-3-01/4-3-01-Seq/4-3-01-Seq-emb1/NA2030 DAT AA746487

seq_documentation_block:

AAZ36687 standard; DNA; 1305 BP.

XX AAZ36687;

XX 13 MAR-2000 (first entry)

XX Nucleotide sequence for GenBank accession number AF019486.

XX Stimulus-regulated nucleic acid; sequence profile; nucleic acid level;
 XX differentially expressed nucleic acid; disease state; cancer;
 KW autoimmune disease; infectious disease; aging; developmental disorder;
 KW proliferative disorder; neurological disorder; toxicity;
 KW treatment resistance; differential expression; drug discovery;
 KW growth factor; epidermal growth factor; radiation; stress; pathogen; ss.

XX Homo sapiens.

XX W09955913-A2.

XX 04-NOV-1999.

XX 27-APR-1999; 9906-US09119.

XX 27-APR-1998; 9805-0083331.

XX 27-AUG-1998; 9805-0098070.

XX 04-FEB-1999; 9905-0118624.

XX (KIMM-) KIMMEL CANCER CENT SIDNEY.

XX McClelland M, Welsh J, Trenkle T;

XX WPI: 2000-086488/07.

XX Measuring expression of low abundance reduced complexity target nucleic

XX acid molecules

XX Disclosure, Fig 14, 187pp, English.

XX AAZ3681 23625 represent nucleotide sequences from stimulus-regulated
 CC nucleic acid molecules. The sequences represent a profile of sequences
 CC which can function as targets in the method of the invention. The
 CC specification describes a method for measuring the level of two or more
 CC nucleic acid molecules in a target. The method comprises contacting a
 CC probe with an arbitrarily or statistically sampled target and detecting
 CC the amount of specific binding of the target to the probe. The methods
 CC can be used to identify differentially expressed nucleic acid molecules
 CC associated with disease states, such as cancer, autoimmune disease,
 CC infectious disease, aging, developmental disorder, proliferative
 CC disorder or neurological disorder. Alternatively the methods can be
 CC used to assess the efficacy of toxicity of or a resistance to a
 CC treatment. Also the methods can be used to determine differential
 CC expression of nucleic acid molecules in response to a stimulus, e.g. a
 CC chemical, drug or growth factor (especially epidermal growth factor),
 CC radiation, stress or a pathogen. The methods can also be used to
 CC determine co-regulated genes that can be potential targets for drug
 CC discovery.

XX Sequence 1305 BP; 304 A; 380 C; 329 G; 292 T; 0 other;

alignment_scores:

Quality: 1526.00

Ratio: 5.392

Percent Similarity: 100.000

Percent Identity: 100.000

alignment_block:

US 09-557-262_4_COPY_21_303 x AAZ36687

Align seq 1/1 to: AAZ36687 from: 1 to: 1305


```

alignment_scores:      Quality: 829.00      Length: 150
                       Ratio: 5.527        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-557-262-4_COPY_21_303 x ABA55377/rev ..

Align seg 1/1 to reverse of: ABA55377 from: 1 to: 452

65 AsnGluValHisPheAspTyrPheGluHisTyrSerHisGlyLeuG1 81
|||||
451 AACGAGGTCCTATTCTTGAATGGAGAGGATTACAGACAGGTTGGG 402

81 YTrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLeuThrValG 98
|||||
401 CTGGTACCTCAGCAAGATGCTTTTCTGCTGCTGCTGCTGCTGCTGCT 452

98 LuLysThrProAlaTyrPheThrSerProLysValProGluArgValTyr 114
|||||
351 AGAAGACCCCTACCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 402

115 SerMetAsnProSerIleArgLeuLeuLeuLeuLeuArgAspProSerG1 131
|||||
301 AATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 352

131 uArgValLeuSerAspTyrThrGluValPheTyrAsnHisMetGlnLysH 148
|||||
251 GAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 302

148 isLysProTyrProSerIleGluPheLeuValArgAspGlyArgLeu 164
|||||
201 ACAAGCCCTACCGCTCCCACTCAGAGTTCCTGCTGCTGCTGCTGCTGCT 152

165 AsnValAspTyrLysAlaLeuAsnArgSerLeuTyrHisValHisMetG1 181
|||||
151 AATGTGATCTATCAAGGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 102

181 nAsnTrpLeuArgPhePheProLeuArgHisIleHisIleValAspGlyA 198
|||||
101 GAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 52

198 spArgLeuIleArgAspProPheProGluIleGlnLysValGluArgPhe 214
|||||
51 ACCGCTCTCATCAGGACCCCTTCCTTGACATCCCAAGCTCGAGGCTTC 2

seq_name: /S105/ycgdata/geneseq/geneseq-n-emb1/NA2001A.DAT:ABA25103

seq_documentation_block:
ID ABA25103 standard; DNA; 452 BP.
XX
AC ABA25103:
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #356q for gene expression analysis in human heart cell sample
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WC200157274-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001W0-NS00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632456.

```

```

PP 21-SEP-2000; 2000US-0234587.
PR 27-SEP-2000; 2000US-0234589.
PR 04-OCT-2000; 2000US-0234593.
XX
XX (MORF-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Harzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX
XX claim 1: SEQ ID No 356q; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX
XX Herein the sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WPI
XX at ftp.wpi.int/pub/published_pat_sequences.
XX
XX Sequence 452 BP; 86 A; 114 C; 155 G; 97 T; 0 other;

```

```

alignment_scores:      Quality: 829.00      Length: 150
                       Ratio: 5.527        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-557-262-4_COPY_21_303 x ABA25103/rev ..

Align seg 1/1 to reverse of: ABA25103 from: 1 to: 452

65 AsnGluValHisPheAspTyrPheGluHisTyrSerHisGlyLeuG1 81
|||||
451 AACGAGGTCCTATTCTTGAATGGAGAGGATTACAGACAGGTTGGG 402

81 YTrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLeuThrValG 98
|||||
401 CTGGTACCTCAGCAAGATGCTTTTCTGCTGCTGCTGCTGCTGCTGCT 352

98 LuLysThrProAlaTyrPheThrSerProLysValProGluArgValTyr 114
|||||
351 AGAAGACCCCTACCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 402

115 SerMetAsnProSerIleArgLeuLeuLeuLeuLeuArgAspProSerG1 131
|||||
301 AATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 352

131 uArgValLeuSerAspTyrThrGluValPheTyrAsnHisMetGlnLysH 148
|||||
251 GAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 302

148 isLysProTyrProSerIleGluPheLeuValArgAspGlyArgLeu 164
|||||
201 ACAAGCCCTACCGCTCCCACTCAGAGTTCCTGCTGCTGCTGCTGCTGCT 152

165 AsnValAspTyrLysAlaLeuAsnArgSerLeuTyrHisValHisMetG1 181
|||||
151 AATGTGATCTATCAAGGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 102

181 nAsnTrpLeuArgPhePheProLeuArgHisIleHisIleValAspGlyA 198
|||||
101 GAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 52

```


CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukemia and myeloma. The present sequence is one of
 CC the probes of the invention
 XX
 SQ Sequence 452 BP; 86 A; 114 C; 155 G; 97 T; 0 other;

alignment_scores:
 Quality: 829.00 length: 150
 Ratio: 5.527 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

us-09-557-262-4_copy_21_303 x AAK29079/rev ..

Align seg 1/1 to reverse of: AAK29079 from: 1 to: 452

```

65 AsnGluValHisPheAspTrpGluHisTyrSerHisGlyLeuG1 81
|||||
451 AAGCAGTCTACCTCTCCACCTGGAGAGAGATTAACAGCCAGGCTTGGG 402

81 YTrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLeuThrValG 98
|||||
401 CTGTATCTTCAAGTAATGAGATTTCTGTGGGAAATTAATTAATTAATTA 352

98 LuLysThrProAlaTyrPheThrSerProLysValProGluArgValTyr 114
|||||
351 AGAAGACCCCGCGTATTTTCAGCTCCCAAGCTGCTTCACGCACTCTAC 302

115 SerMetAsnProSerIleArgLeuLeuLeuLeuLeuLeuLeuLeuLeu 131
|||||
301 ACATCAACCCGCTCCATCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 252

131 uArgValLeuSerAspTyrThrGlnValPheTyrAsnHisMetGlnTyr 148
|||||
251 GCGGTGTATTTATTTATTAATTAATTAATTAATTAATTAATTAATTA 202

148 isLysProTyrProSerIleGluGluPheLeuValArgSpGlyArgLeu 164
|||||
201 ACAAGCCCTACCGCTCCATCTGAGGAGTTTCTGCTGCGGATGCGAGGCTC 152

165 AsnValAspTyrLysAlaLeuAsnArgSerLeuTyrHisValHisMetG1 181
|||||
151 AATCTGACTTACAGGCTCTTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCT 102

181 nasnTrpLeuArgPhePheProLeuArgHisIleHisIleValAspGly 198
|||||
101 GAACGTGTATTTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 52

198 spArgLeuIleArgAspProPheProGluIleGlnLysValGluArgPhe 214
|||||
51 ACCGCTCTATCAAGGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2

```

seq_name: /SID55/gcdata/geneseq/geneseq-emb1/NA2001A.DAT.AAI35025

seq_documentation_block:

ID AAI35025 standard; DNA: 452 BP.

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX AC

XX

PF

XX

XX

PR

PR

PR

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

PP

30-JAN-2001; 2001WO-0500663.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-APR-2000; 2000US-0612466.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0619359.

04-OCT-2000; 2000CB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG. Hanzel DK. Chen W. Pank DR.

WPI: 2001-488897/53.

Human genome derived single exon nucleic acid probes useful for

analyzing gene expression in human placenta.

Claim 25; SEQ ID NO 3711; 654pp; English.

The present invention relates to single exon nucleic acid probes (SNP)

The present sequence is one such probe. The probes are useful for

producing a microarray for predicting, measuring and displaying gene

expression in samples derived from human placenta. The probes are useful

for antenatal diagnosis of human genetic disorders.

Sequence 452 BP; 86 A; 114 C; 155 G; 97 T; 0 other;

alignment_scores:

Quality: 829.00 length: 150

Ratio: 5.527 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

us-09-557-262-4_copy_21_303 x AAI35025/rev ..

Align seg 1/1 to reverse of: AAI35025 from: 1 to: 452

65 AsnGluValHisPheAspTrpGluHisTyrSerHisGlyLeuG1 81

|||||

451 AAGCAGTCTACCTCTCCACCTGGAGAGAGATTAACAGCCAGGCTTGGG 402

81 YTrpTyrLeuSerGlnMetProPheSerTrpProHisGlnLeuThrValG 98

|||||

401 CTGTATCTTCAAGTAATGAGATTTCTGTGGGAAATTAATTAATTAATTA 352

98 LuLysThrProAlaTyrPheThrSerProLysValProGluArgValTyr 114

|||||

351 AGAAGACCCCGCGTATTTTCAGCTCCCAAGCTGCTTCACGCACTCTAC 302

115 SerMetAsnProSerIleArgLeuLeuLeuLeuLeuLeuLeuLeuLeu 131

|||||

301 ACATCAACCCGCTCCATCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 252

131 uArgValLeuSerAspTyrThrGlnValPheTyrAsnHisMetGlnTyr 148

|||||

251 GCGGTGTATTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 202

148 isLysProTyrProSerIleGluGluPheLeuValArgSpGlyArgLeu 164

|||||

201 ACAAGCCCTACCGCTCCATCTGAGGAGTTTCTGCTGCGGATGCGAGGCTC 152

165 AsnValAspTyrLysAlaLeuAsnArgSerLeuTyrHisValHisMetG1 181

|||||

151 AATCTGACTTACAGGCTCTTCAAGCTCAAGCTCAAGCTCAAGCTCAAGCT 102

181 nasnTrpLeuArgPhePheProLeuArgHisIleHisIleValAspGly 198

|||||

101 GAACGTGTATTTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 52

198 SPARGUEUTICARGLASPRIPIEPROGLUTILEGLUTYSLVALGLUARGPHE 214
|||||
51 ACCGGGTCATCGAGGAGGCGTTCCTGAGATCGAAAAAGTGGAGGCTTC 2

```

seq_name: gb_est1:AA407647

seq_documentation_block:
LOCUS      AA407647          362 bp    mRNA       linear     EST 26-AUG-1998
DEFINITION Mouse 7.5 dpc embryo ectoplacental cone cDNA library Mus musculus cDNA clone p0004A04 5', mRNA sequence
ACCESSION  AA407647
VERSION    AA407647.1   GI:2065848
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 362)
AUTHORS   Ko,M.S.H., Threat,T.A., Horton,J.H., Wang,X., Cui,Y., Wang,X., Pryor,E., Paris,J., Wells-Smith,J., Fujiwara,H., Yotsumoto,S. and Nakashima,H.
TITLE     Systematic analyses of mouse genes expressed in embryo implantation site
JOURNAL    Unpublished (1997)
COMMENT    Other ESTs: EST00932
Contact: Koshida
Wayne State University
5047 Gullen Mall, Detroit, MI 48202
Tel: 3135776708
Fax: 3135776200
Email: msk@omb.biosci.wayne.edu
Seq primer: M13 Reverse.
Location/Qualifiers
1..362
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="ATCC (inhost):1363251"
/db_xref="taxon:10090"
/clone="C0004X09"
/library="Mouse 7.5 dpc embryo ectoplacental cone cDNA library"
/sex="unknown"
/tissue_type="ectoplacental cone"
/dev_stage="embryonic day 7.5 postconception"
/lab_host="DH10B"
/notes="Organ: embryo; Vector: pSPcPTT (Life Technologies); Site_1: SalI; Site_2: NotI; Total RNAs were extracted from ectoplacental cone of 7.5-dpc embryos. The double-stranded cDNA was synthesized from total RNAs with an Oligo(dT) primer. The library was constructed by Minoru S. H. Ko."

BASE COUNT      105 a      83 c      77 g      97 t

alignment_scores:
Quality: 155.00      Length: 27
Ratio: 5.741        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-557-262-2_COPY_250_276 x AA407647 ..
Align seq 1/1 to: AA407647 from: 1 to: 362
1 PheTyRPheAsnlysthrlysGlyPheTyrcysLeuArgaspSerGlyly 17
|||||
12 TTCTACTTTTAACAACCAAGCGCTCTCTATTGGTGCGCATGTGCATAA 61

17 sAspArgCysLeuHislnlsrLyserlysglyArg 27
|||||
62 GCACCCTGCTTACACGAGTCCTCAAGGCCCG 92

seq_name: gb_est2:W62484
seq_documentation_block:
LOCUS      W62484
DEFINITION md72c12.r1 Soares mouse embryo NM0E13.5 14.5 Mus musculus cDNA clone IMAGE:373942 5' similar to PIR:A49733 A49733 glycosaminoglycan N-acetylglucosaminyl N-deacetylase ;, mRNA

seq_name: gb_est1:AA407647
seq_documentation_block:
LOCUS      AA407647          362 bp    mRNA       linear     EST 26-AUG-1998
DEFINITION Mouse 7.5 dpc embryo ectoplacental cone cDNA library Mus musculus cDNA clone p0004A04 5', mRNA sequence
ACCESSION  AA407647
VERSION    AA407647.1   GI:2065848
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 362)
AUTHORS   Ko,M.S.H., Threat,T.A., Horton,J.H., Wang,X., Cui,Y., Wang,X., Pryor,E., Paris,J., Wells-Smith,J., Fujiwara,H., Yotsumoto,S. and Nakashima,H.
TITLE     Systematic analyses of mouse genes expressed in embryo implantation site
JOURNAL    Unpublished (1997)
COMMENT    Other ESTs: EST00932
Contact: Koshida
Wayne State University
5047 Gullen Mall, Detroit, MI 48202
Tel: 3135776708
Fax: 3135776200
Email: msk@omb.biosci.wayne.edu
Seq primer: M13 Reverse.
Location/Qualifiers
1..362
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="ATCC (inhost):1363251"
/db_xref="taxon:10090"
/clone="C0004X09"
/library="Mouse 7.5 dpc embryo ectoplacental cone cDNA library"
/sex="unknown"
/tissue_type="ectoplacental cone"
/dev_stage="embryonic day 7.5 postconception"
/lab_host="DH10B"
/notes="Organ: embryo; Vector: pSPcPTT (Life Technologies); Site_1: SalI; Site_2: NotI; Total RNAs were extracted from ectoplacental cone of 7.5-dpc embryos. The double-stranded cDNA was synthesized from total RNAs with an Oligo(dT) primer. The library was constructed by Minoru S. H. Ko."

BASE COUNT      105 a      83 c      77 g      97 t

alignment_scores:
Quality: 155.00      Length: 27
Ratio: 5.741        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-557-262-2_COPY_250_276 x AA407647 ..
Align seq 1/1 to: AA407647 from: 1 to: 362
1 PheTyRPheAsnlysthrlysGlyPheTyrcysLeuArgaspSerGlyly 17
|||||
12 TTCTACTTTTAACAACCAAGCGCTCTCTATTGGTGCGCATGTGCATAA 61

17 sAspArgCysLeuHislnlsrLyserlysglyArg 27
|||||
62 GCACCCTGCTTACACGAGTCCTCAAGGCCCG 92

seq_name: gb_est2:W62484
seq_documentation_block:
LOCUS      W62484
DEFINITION md72c12.r1 Soares mouse embryo NM0E13.5 14.5 Mus musculus cDNA clone IMAGE:373942 5' similar to PIR:A49733 A49733 glycosaminoglycan N-acetylglucosaminyl N-deacetylase ;, mRNA

```



seq_name. gb_est2:BM122797

seq_documentation_block:

LOCUS BM122797 492 bp. mRNA linear EST 01-FEB-2002

DEFINITION L0515C10-3 NIA Mouse Newborn Heart cDNA Library Mus musculus cDNA

clone L0515C10 3', mRNA sequence.

ACCESSION BM122797

VERSION BM122797

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 492)

AUTHORS Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Pantano, S., Lim, M.K. and Ko, M.S.H.

TITLE Systematic Analyses of NIA Mouse Newborn Heart cDNA Library

JOURNAL Unpublished (2001)

COMMENT Contact: Dawood B. Dudekula Laboratory of Genetics National Institute on Aging/National Institutes of Health 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6829, USA Email: cdnaelgsun.grc.nia.nih.gov Plate: L0515 row: C column: 10 Seq primer: -21M13 Forward High quality sequence stop. 492 POLYA-Yes.

FEATURES

Location/Qualifiers

1. 492

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="niaEST:L0515C10-3"

/db_xref="taxon:10090"

/clone="L0515C10"

/clone_lib="NIA Mouse Newborn Heart cDNA Library"

/tissue_type="Newborn Heart"

/dev_stage="Newborn"

/lab_host="DH10B"

/note="organ: heart; Vector: pSPORT1 (Invitrogen); Site: 1: SalI, Site: 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsn.grc.nia.nih.gov/GUNA). Double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen: 5' pACTAGTCTAGATCGGCGGCCGCCCTTTT-3'] from 24.9 microgram of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal3 (Ref. Development 127:1737-1749 (2000) [PMID: 10725249]), purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were digested with SalI and NotI enzymes, and cloned into SalI and NotI site of pSPORT1 plasmid vector. The pSPORT1 E. coli host was transformed with ligation mixture by the chemical method. The average insert size is about 1.8 kb. The library was constructed by Yulan Piao (NIA)."

BASE COUNT 157 a 102 c 91 g 142 t

ORIGIN

alignment_scores:

Quality: 155.00 Length: 27

Ratio: 5.741 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-557-262-2_copy_250_276 x BM122797/rev

Align seg 1/1 to reverse of. BM122797 from. 1 to: 492

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGly 17

|||||

488 TTCTACTTTTAAACAAACCAAGGCTTCTACTGCTGCGGACAGTGCA 439

sequence.

W62484

VERSION W62484.1 GI:1369243

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 490)

AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.H.

TITLE The WashU-HHMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:235374

Seq primer: ETPrimer

High quality sequence stop: 344.

Location/Qualifiers

1. 490

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:373942"

/clone_lib="Soares mouse embryo NBMEL3.5 14.5"

/sex="unknown"

/tissue_type="embryo"

/dev_stage="13.5-14.5dpc total fetus"

/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site: 1: Not I; Site: 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGCGCGCGGAAATTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 127 a 129 c 112 g 122 t

ORIGIN

alignment_scores:

Quality: 155.00 Length: 27

Ratio: 5.741 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-557-262-2_COPY_250_276 x W62484

Align seg 1/1 to: W62484 from: 1 to: 490

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGly 17

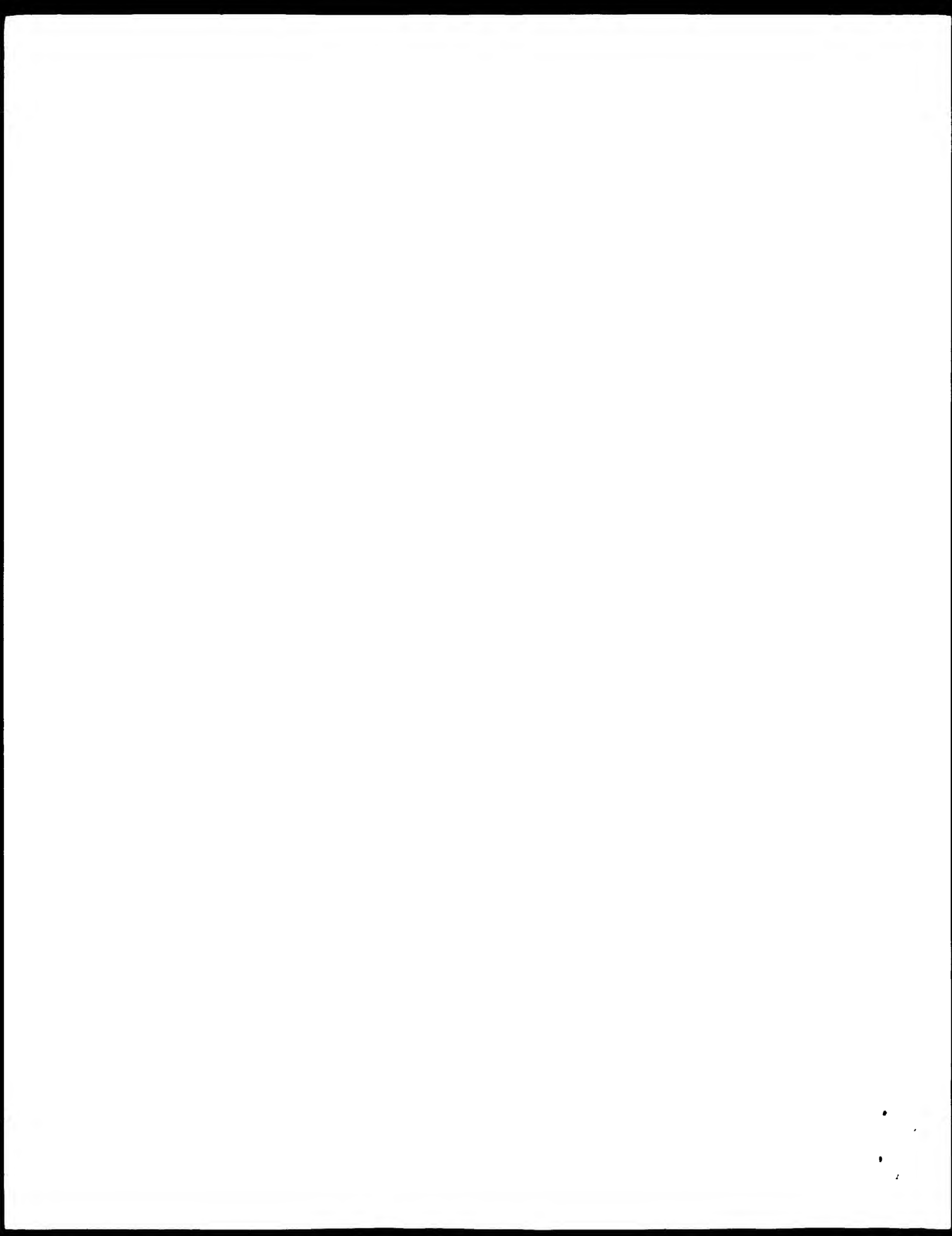
|||||

196 TTCTACTTTTAAACAAACCAAGGCTTCTACTGCTGCGGACAGTGCA 245

17 sAspArgCysLeuHisGluSerLysGlyArg 27

|||||

246 GGACCGCTGCTTACAGGCTTCAAGGCGCGG 276



[illegible]

Percent sum of squares	Percent identity
100.000	100.000

alignment_block:

US-09-557-262-2_copy_250_276 x AAX37245

Align seq 1/1 to: AAX37245 from: 1 to: 1685

1 PheTyRPhaAsnLysThrLysGlyPheTyRcysLeuArgAspSerGlyLy 17
|||||
1070 TTCTACTTTTAAACAAACCAAGGGGCTTCTACGCTGCGGGAATGATGAGTAA 1119

17 sAspArgcysLeuHisGluSerLysGlyArg 27
|||||
1120 GCAACCGCTGCTTACACGACGTCCTCAAAAGCGCGG 1150

seq_name: /SID55/genedata/geneseq/emb1/NA2001A.DAT:ABA50047

seq_documentation_block:

ID ABA50047 standard: DNA: 924 BP

XX ABA50047;

XX AC

XX 01-FEB-2002 (first entry)

XX DE Human breast cell single exon nucleic acid probe #8742.

XX KW Human; microarray, single exon probe, gene expression, breast,

XX KW disease; cancer; ss.

XX OS Homo sapiens.

XX PN W0200157271-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001W0-US00662.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 01-APR-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000US-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DP.

XX DR WPI; 2001-496433/54

XX CC New spatially addressable set of single exon nucleic acid probes, useful for measuring gene expression in sample derived from human breast, comprises number of single exon nucleic acid probes -

XX PS Claim 4; SEQ ID NO 8742; 327bp + sequence listing; English.

XX CC The invention relates to a spatially addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human breast and BT 474 cells. The method involves contacting the probes with a collection of detectably labelled nucleic acids derived from mRNA of human breast, and then measuring the label bound to each probe of the microarray. The probes are useful for verifying the expression of regions of genomic DNA predicted to encode proteins. They are useful for gene discovery, and for determining predisposition and/or prognosing breast disease. Gene expression analysis is useful for assessing the toxicity of chemical agents on cells. The microarray of this invention presents a far greater diversity of probes for measuring gene expression, with far less bias than expressed sequence tag microarrays. The method is suitable for rapid production of functional information from genomic sequence. The present sequence is a single exon nucleic acid probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/pub/published_pat_sequences

XX

SQ Sequence 924 BP; 169 A, 252 C, 310 G, 193 T, 0 other;

alignment_scores:

Quality: 152.00 Length: 27
Ratio: 5.630 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:

US-09-557-262-2_copy_250_276 x ABA50047/rev

Align seq 1/1 to reverse of: ABA50047 from: 1 to: 924

1 PheTyRPhaAsnLysThrLysGlyPheTyRcysLeuArgAspSerGlyLy 17
|||||
170 TTCTACTTTTAAACAAACCAAGGGGCTTCTACGCTGCGGGAATGATGAGTAA 211
17 sAspArgcysLeuHisGluSerLysGlyArg 27
|||||
120 GCAACCGCTGCTTACACGACGTCCTCAAAAGCGCGG 90

seq_name: /SID55/genedata/geneseq/emb1/NA2001A.DAT:ABA67978

seq_documentation_block:

ID ABA67978 standard: DNA: 924 BP

XX ABA67978;

XX 01-FEB-2002 (first entry)

XX DE Human foetal liver single exon nucleic acid probe #16283.

XX KW Human, foetal liver, gene expression; single exon nucleic acid probe; ss.

XX OS Homo sapiens.

XX PN W0200157277 A2.

XX PR 09-AUG-2001.

XX PR 30-JAN-2001; 2001W0-US00663.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 01-APR-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000US-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DP;

XX DR WPI; 2001-483447/52.

XX CC Human genome-derived single exon nucleic acid probes useful for

XX FT analyzing gene expression in human fetal liver.

XX PS Claim 4; SEQ ID NO 16283; 639bp + sequence listing; English.

XX CC The invention relates to a single exon nucleic acid probe for

XX CC measuring human gene expression in a sample derived from human foetal

XX CC liver. The single exon nucleic acid probes may be used for predicting,

XX CC measuring and displaying gene expression in samples derived from human

XX CC foetal liver. The present sequence is a single exon nucleic acid

XX CC probe of the invention.

XX CC Note: The sequence data for this patent did not form part of the

XX CC printed specification, but was obtained in electronic format directly

XX CC from WIPO at http://wipo.int/pub/published_pat_sequences.

XX SQ Sequence 924 BP; 169 A, 252 C, 310 G, 193 T, 0 other;

```

alignment_scores:
  Quality: 152.00      Length: 27
  Ratio: 5.630         Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:
US-09-557-262-2_COPY_250_276 x ABA67978/rev
Align seq 1/1 to reverse of: ABA67978 from: 1 to: 924

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyLy 17
170 TTCTACTTTTAAACAAACCAAGGGCTTTTACTGCTTGGTGGACACGGCG 121

17 sAspArgCysLeuHisGluSerLysGlyArg 27
120 GGACCTCTCTTACATGACTCCAAAGCGCG 90

seq_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:ABA35017
seq_documentation_block:
ID ABA35017 standard; DNA: 924 BP.
XX
AC ABA35017;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #13483 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000666.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 4; SEQ ID No 13483; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

```

```

CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 924 BP; 169 A; 252 C; 310 G; 193 T; 0 other;
XX
alignment_scores:
  Quality: 152.00      Length: 27
  Ratio: 5.630         Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:
US-09-557-262-2_COPY_250_276 x ABA35017/rev
Align seq 1/1 to reverse of: ABA35017 from: 1 to: 924

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyLy 17
170 TTCTACTTTTAAACAAACCAAGGGCTTTTACTGCTTGGTGGACACGGCG 121

17 sAspArgCysLeuHisGluSerLysGlyArg 27
120 GGACCTCTCTTACATGACTCCAAAGCGCG 90

seq_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAK16371
seq_documentation_block:
ID AAK16371 standard; DNA: 924 BP.
XX
AC AAK16371;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 16362.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000667.
XX
PR 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
XX
PR 30-JUN-2000; 2000US-0608408.
XX
PR 03-AUG-2000; 2000US-0632366.
XX
PR 21-SEP-2000; 2000US-0234687.
XX
PR 27-SEP-2000; 2000US-0236359.
XX
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR.
XX
WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 16362; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX

```


US-09-557-262-2_COPY_250_276 x AAI48186/rev ..

Align seg 1/1 to reverse of AAT48186 from 1 to 924

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArqAspScrGlyLy 17
 |||||||
 170 TTCTACTTTTAAACAAACCAAGGCTTTTACTGCTCGCGGACACCGGCCG 121
 17 sAspArqCysLeuLisGluSerLysGlyArg 27
 :|||||
 120 GGACGCTGCTTATCATGATGATGAAAGGCGG 90

seq_name: /SID55/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AA37246

seq_documentation_block:

ID AAX37246 standard; DNA; 1305 BP.

AC AAX37246;

DT 20-JUL-1999 (first entry)

DE Human 3-OST-1 encoding DNA.

KW Heparan sulfate D-glucosaminyl 3-O-sulfotransferase; anticoagulant;
 saccharide; glycosaminoglycan; proteoglycan; antithrombin; 3-OST;
 KW heparan sulfate; thrombotic disorder; deep vein thrombosis;
 KW pulmonary embolism; coagulation enzyme inactivation; ss

XX Homo sapiens.

XX W09922005-A2

XX 06-MAY-1999.

XX 23-OCT-1998; 98WO-0522597.

XX 31-OCT-1997; 97US-0065437.

XX 24-OCT-1997; 97US-0062762.

PA (MASI) MASSACHUSETTS INST TECHNOLOGY.

PI Fritze LMS, Liu J, Rosenberg RD, Schwartz JJ, Shworak NW,
 Zhang L;

XX WPI; 1999-312968/26.

DR P-PSDB; AAY17063.

XX Heparan sulfate D-glucosaminyl 3-O-sulfotransferase and related
 PT polynucleotides

XX Claim 5; Page 74-75; 95pp; English.

XX The invention relates to nucleic acid molecules (AAX37245-X37250)
 CC encoding heparan sulfate D-glucosaminyl 3-O-sulfotransferases (3-OSTs).
 CC The 3-OST proteins can be used for 3-O-sulfating saccharide residues
 CC within a preparation of glycosaminoglycan or proteoglycan
 CC polysaccharides. 3-OST-1 can be used for enriching the
 CC antithrombin-binding fraction in a preparation of heparan sulfates (HS).
 CC 3-OST-1 can also convert HS proteoglycan anticoagulant precursor to the
 CC HS anticoagulant (HSact). The 3-OSTs (optionally lacking enzymatic
 CC function) can be used to determine partial sequence information for
 CC complex polysaccharides. The 3-OST proteins, genes and antibodies are
 CC also useful for diagnosis of disorders involving HS biosynthesis. 3-OSTs
 CC are especially used to generate anticoagulant pentasaccharides, which may
 CC be used to treat thrombotic disorders such as deep vein thrombosis and
 CC pulmonary embolism. Coagulation enzyme inactivation by antithrombin is
 CC enhanced by complexing of antithrombin with endothelial cell surface HS
 CC proteoglycans. This is responsible for the non-thrombogenic properties
 CC of blood vessels. The present sequence represents a human 3-OST-1
 CC encoding DNA.

XX Sequence 1305 BP; 304 A; 380 C; 329 G; 292 T; 0 other;

alignment_scores:

Quality: 152.00 Length: 27
 Ratio: 5.630 Gaps: 0
 Percent similarity: 100.000 Percent identity: 96.296

alignment_block:

US-09-557-262_2_COPY_250_276 x AAX37246 ..

Align seg 1/1 to: AAX37246 from: 1 to: 1305

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArqAspScrGlyLy 17
 |||||||
 854 TTCTACTTTTAAACAAACCAAGGCTTTTACTGCTCGCGGACACCGGCCG 903
 17 sAspArqCysLeuLisGluSerLysGlyArg 27
 :|||||
 904 GGACGCTGCTTATCATGATGATGAAAGGCGG 934

seq_name: /SID55/gcgdata/geneseq/gcgseqn-emb1/NA2068.DAT:AA236687

seq_documentation_block:

ID AAX36687 standard; DNA; 1305 BP.

AC AAX36687;

DT 13-MAR-2000 (first entry)

DE Nucleotide sequence for GenBank accession number AF019386.

XX Stimulus regulated nucleic acid, sequence profile, nucleic acid level;
 KW differentially expressed nucleic acid; disease state; cancer;
 KW autoimmune disease; infectious disease; aging; developmental disorder;
 KW proliferative disorder; neurological disorder; toxicity;
 KW treatment resistance; differential expression; drug discovery;
 KW growth factor; epidermal growth factor; radiation; stress; pathogen; ss.
 XX Homo sapiens.

XX W09955913-A2.

XX 04-NOV-1999.

XX 27-APR-1999; 99WO-US09119.

XX 27-APR-1998; 98US-0083331.

XX 27-APR-1998; 98US-0083331.

XX 04-FEB-1999; 99US-0118624.

XX (KIMM-) KIMMEL CANCER CENT SIDNEY.

XX McClelland M, Welsh J, Trenkle T;

XX WPI; 2000-086188/07.

XX Measuring expression of low abundance reduced complexity target nucleic
 PT acid molecules

XX Disclosure; Fig 14; 187pp; English.

XX AAX36681-X36725 represent nucleotide sequences from stimulus-regulated
 CC nucleic acid molecules. The sequences represent a profile of sequences
 CC which can function as targets in the method of the invention. The
 CC specific information describes a method for measuring the level of two or more
 CC nucleic acid molecules in a target. The method comprises contacting a
 CC probe with an arbitrarily or statistically sampled target and detecting
 CC the amount of specific binding of the target to the probe. The methods
 CC can be used to identify differentially expressed nucleic acid molecules
 CC associated with disease states, such as cancer, autoimmune disease,
 CC infectious disease, aging, developmental disorder, proliferative
 CC disorder or neurological disorder. Alternatively the methods can be
 CC used to assess the efficacy of toxicity of or a resistance to a
 CC treatment. Also the methods can be used to determine differential

CC expression of nucleic acid molecules in response to a stimulus, e.g. a
 CC chemical, drug or growth factor (especially epidermal growth factor),
 CC radiation, stress or a pathogen. The methods can also be used to
 CC determine co-regulated genes that can be potential targets for drug
 CC discovery.

XX
 SQ Sequence 1305 BP; 304 A; 380 C; 329 G; 292 T; 0 other;

alignment_scores:
 Quality: 152.00 Length: 27
 Ratio: 5.630 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:
 US-09-557-262-2_COPY_250_276 x AA236687

Align seq 1/1 to: AA236687 from: 1 to: 1305

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyly 17
 |||||
 854 TTCTACTTTTAA/AAAAACCA/AGG/TTTTAT/ATG/TAT/AGGAA/AT/AGGAG 903

17 sAspArgCysLeuHisGluSerLysGlyArg 27

|||||
 904 GCACCGCTGCTTACATGAGTCCAAAGCCCG 934

seq_name: /SIDS5/gcgdata/gene-seq/geneseq-nmb1/NA2001R.1A1.ABL26623

seq_documentation_block:

ID: ABL26623 standard; DNA; 897 BP.

XX AC ABL26623;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 31342.

XX KW Drosophila; developmental biology; cell signalling, insecticide,
 XX pharmaceutical; gene; ds.

XX OS Drosophila melanogaster.

XX PN W0200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001W0-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX PS Claim 1: SEQ ID NO 31342; 21pf - Sequence Listing, English.

XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABH57737-ABH72072).

XX CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 897 BP; 183 A; 233 C; 295 G; 186 T; 0 other;

alignment_scores:
 Quality: 114.00 Length: 27
 Ratio: 4.560 Gaps: 0
 Percent Similarity: 92.593 Percent Identity: 74.074

alignment_block:
 US-09-557-262-2_COPY_250_276 x ABL26623

Align seq 1/1 to: ABL26623 from: 1 to: 897

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyly 17
 |||||
 703 TTTTACTTCAATGAGACCAAGGCTTCTACTGTCGCTTATGACACCG 752

17 sAspArgCysLeuHisGluSerLysGlyArg 27

|||||
 753 GATGCTTATTTGAGAGAGAAAGGAAAGG 783

seq_name: /SIDS5/gcgdata/gene-seq/geneseq-nmb1/NA2001R.1A1.ABL26622

seq_documentation_block:

ID: ABL26622 standard; DNA; 348; BP.

XX AC ABL26622;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 31349.

XX KW Drosophila; developmental biology; cell signalling, insecticide,
 XX pharmaceutical; gene; ds.

XX OS Drosophila melanogaster.

XX PN W0200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001W0-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX PS Claim 1: SEQ ID NO 31339; 21pp - Sequence Listing, English.

XX CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABH57737-ABH72072).

XX CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

CC complex polysaccharides. The 3-OST proteins, genes and antibodies are
 CC also useful for diagnosis of disorders involving HS biosynthesis. 3-OSTs
 CC are especially used to generate anticoagulant pentasaccharides, which may
 CC be used to treat thrombotic disorders such as deep vein thrombosis and
 CC pulmonary embolism. Coagulation enzyme inactivation by antithrombin is
 CC enhanced by complexing of antithrombin with endothelial cell surface HS
 CC proteoglycans. This is responsible for the non-thrombotic properties
 CC of blood vessels. The present sequence represents a human 3-OST-4
 CC encoding DNA.

XX
 SQ Sequence 3658 BP; 771 A, 1085 C, 1167 G; 695 I, 0 other;

alignment_scores:
 Quality: 97.50 Length: 30
 Ratio: 4.239 Gaps: 1
 Percent Similarity: 76.667 Percent Identity: 66.667

alignment_block:
 US-09-557-262-2_COPY_250_276 x AAX37250 ..

Align seg 1/1 to: AAX37250 from: 1 to: 3658

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArg.....As 14
 |||||
 2002 TCTATTTCACACAGACGACAAAGGATTCCTTGCTTGAAAGAAAGAAAGA 2051
 |||||
 14 pSerGlyLysAspArgCysLeuHisGluSerLysGlyArg 27
 |||||
 2052 CAGCAGTCGCGCGAGAGGAGTTAGGAAAGAAAGAAAGG 2091

seq_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT: AAX37247

seq_documentation_block:

ID AAX37247 standard; DNA, 1951 BP.

XX AAX37247,

DE 20-JUL-1999 (first entry)

DE Human 3-OST-2 encoding DNA.

XX Heparan sulfate D-glucosaminyl 3-O-sulfotransferase; anticoagulant;
 KW saccharide; glycosaminoglycan; proteoglycan; antithrombin; 3-OST;
 KW heparan sulfate; thrombotic disorder; deep vein thrombosis;
 KW pulmonary embolism; coagulation enzyme inactivation; ss.

XX Homo sapiens.

XX W09922005-A2.

XX 06-MAY-1999.

XX 23-OCT-1997; W0991-0522597.

XX 31-OCT-1997; 97NS-0065437.

PR 24-OCT-1997; 97US-0062762.

XX (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX Fritze JMS, Liu J, Fosenberg R, Schwartz JJ, Shwartz NW.

PI Zhang L;

XX WPL; 1999-312868/26.

DR P-PSDB; AAY17064.

XX Heparan sulfate D-glucosaminyl 3-O-sulfotransferase and related
 PT polynucleotides

XX Claim 5; Page 76-78, 95pp; English.

XX The invention relates to nucleic acid molecules (AAX37245-X7250)

CC encoding heparan sulfate D-glucosaminyl 3-O-sulfotransferases (3-OSTs)

CC The 3-OST proteins can be used for 3-O-sulfating saccharide residues
 CC within a preparation of glycosaminoglycan or proteoglycan
 CC polysaccharides. 3-OST-1 can be used for enriching the
 CC antithrombin-binding fraction in a preparation of heparan sulfates (HS).
 CC 3-OST-1 can also convert HS proteoglycan anticoagulant precursor to the
 CC HS anticoagulant (disac). The 3-OSTs (optionally lacking enzymatic
 CC function) can be used to determine partial sequence information for
 CC complex polysaccharides. The 3-OST proteins, genes and antibodies are
 CC also useful for diagnosis of disorders involving HS biosynthesis. 3-OSTs
 CC are especially used to generate anticoagulant pentasaccharides, which may
 CC be used to treat thrombotic disorders such as deep vein thrombosis and
 CC pulmonary embolism. Coagulation enzyme inactivation by antithrombin is
 CC enhanced by complexing of antithrombin with endothelial cell surface HS
 CC proteoglycans. This is responsible for the non-thrombotic properties
 CC of blood vessels. The present sequence represents a human 3-OST-2
 CC encoding DNA.

XX Sequence 1951 BP; 424 A; 603 C; 490 G; 434 T; 0 other;

alignment_scores:
 Quality: 87.50 Length: 30
 Ratio: 3.804 Gaps: 1
 Percent Similarity: 76.667 Percent Identity: 60.000

alignment_block:

US-09-557-262-2_COPY_250_276 x AAX37247 ..

Align seg 1/1 to: AAX37247 from: 1 to: 1951

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyLy 17

|||||

979 TCTATTTCACACAGACGACAAAGGATTCCTTGCTTGAAAGAAAGAAAGA 1028

17 Asp.....ArgCysLeuHisGluSerLysGlyArg 27

|||||

1029 GAGCAGTCGCGCGAGAGGAGTTAGGAAAGAAAGAAAGG 1068

seq_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA2891A.DAT: AAS21347

seq_documentation_block:

ID AAS21347 standard; cDNA; 2845 BP.

XX AAS21347;

XX 24-OCT-2001 (first entry)

XX Human cDNA sequence encoding for pPro504 polypeptide.

XX Human secretory and transmembrane, PRO, mammalian, cancer, lung,

KW breast, prostate, cervical; tumour necrosis factor-alpha; TNF-alpha;

KW cartilage, ear, proliferation, glucose, free fatty acid, skeletal muscle;

KW adipocyte; A-peptide; factor VIIA; gene therapy; ss.

XX Homo sapiens.

XX WC200140466-A2.

XX 07-JUN-2001.

XX 01-DEC-2000; 2000W0-05342678.

XX 01-DEC-1999; 99W0-US28301.

PR 01-DEC-1999; 99W0-US28634.

PR 02-DEC-1999; 99W0-US28551.

PR 02-DEC-1999; 99W0-US28564.

PR 02-DEC-1999; 99W0-US28565.

PR 09-DEC-1999; 99NS-0170262.

PR 16-DEC-1999; 99W0-US30095.

PR 20-DEC-1999; 99W0-US30911.

PR 30-DEC-1999; 99W0-US30999.

PR 30-DEC-1999; 99W0-US31243.

PR 06-JAN-2000; 2000W0-US00377.

1427 GAGGTTCTTCTGTTGATGTTGGGTAATTAATAAAGAGAGA 1466

```

PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 18-FEB-2000; 2000WO-US04342.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 01-MAR-2000; 2000WO-US05004.
PR 01-MAR-2000; 2000WO-US05601.
PR 20-MAR-2000; 2000WO-US07377.
PR 21-MAR-2000; 2000WO-US07532.
PR 30-MAR-2000; 2000WO-US08439.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 10-NOV-2000; 2000WO-US30873.
XX
PA (GETH ) GENENTECH INC.
XX
PI Baker KP, Beresini M, Deforqe L, Desnoyers L, Filvaroli E, Gao W;
PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WJ, Zhang Z;
XX
DR WPI: 2001-408281/43.
DR P-PSDB; AAU12275.
XX
PT Isolated, secretory and transmembrane PRO polypeptide used to detect
PT other PRO polypeptides, link bioactive molecules to cells expressing
PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
PT lung, breast, prostate, cervical .
XX
PS Claim 3; Fig 207; 813pp; English.
XX
XX AAS21244 AAS21518 encode for novel human secretory and transmembrane
CC PRO polypeptides. The PRO polypeptides are useful to detect other
CC PRO polypeptides, to link bioactive molecules to cells expressing
CC PRO polypeptides, to modulate biological activities of cells expressing
CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC polypeptide expression in a cell sample to that in a control sample.
CC Some of the 275 sequences are also useful to stimulate the release of
CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
CC proliferation or differentiation of chondrocytes, the proliferation or
CC gene expression in pericyte cells, the release of proteoglycans from
CC cartilage, the proliferation of inner ear utricular supporting cells or
CC of T-lymphocytes, the release of a cytokine from peripheral blood
CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC skeletal muscle cells or by adipocytes, or inhibit binding of A-peptide
CC to factor VIIA. The PRO polypeptides can be used in assays to identify
CC molecules involved in binding interactions. The polynucleotides encoding
CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC transgenic or knock out animals and can be used in gene therapy.
XX
SQ Sequence 2845 BP; 673 A; 839 C; 747 G; 586 T; 0 other;

```

```

alignment_scores:
  Quality: 87.50      Length: 30
             Ratio: 3.804      Gaps: 1
Percent Similarity: 76.667   Percent Identity: 60.000

```

```

alignment_block:
US-09-557-262_2_copy_250_276 x AAS21347

```

```

Align seq 1/1 to: AAS21347 from: 1 to: 2845

```

```

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuAraAspSerGlyLy 17
|||||
1377 TCTATTTCACACACACAGAGATCCCTTGCITCAAAAAACAGAAATC 1426
|||||
17 sAsp.....ArgCysLeuHisGluSerLysGlyArg 27
|||||

```





```

DEFINITION      W06c11.x1 NCL_CGAP_G06 Homo sapiens cDNA clone IMAGE:2480756 3'
                  similar to TR014792 014792 HEPARAN SULFATE 3-O-SULFOTRANSFERASE-1
PRECURSORS: 17 mRNA sequence.
ACCESSION       AI971626
VERSION         AI971626.1 GI:5768452
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE       1 (bases 1 to 651)
AUTHORS        NCL_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL        Tumor Gene Index
COMMENT        Unpublished (1997)
                Contact: Robert Strausberg, Ph.D.
                Email: cgapbs@mail.nih.gov
                Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
                R. Emmert-Buck, M.D., Ph.D.
                cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
                Bonaldo, Ph.D.
                cDNA Library Arrayed by: Greg Lennon, Ph.D.
                DNA Sequencing by: Washington University Genome Sequencing Center
                Clone distribution: NCI-CGAP clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                www.bio.llnl.gov/bbrp/image/image.html
                Insert Length: 1020 Std Error: 0.00
                Seq primer: -400P from Gibco
                High quality sequence stop: 456.
FEATURES        source
                  1..651
                  location/Qualifiers
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone IMAGE:2480756"
                    /clone_lib="NCL_CGAP_G06"
                    /tissue_type="pooled germ cell tumors"
                    /lab_host="H1h10"
                    /note="Vector: pTZ19-pac (Pharmacia) with a modified
                    polylinker; Site_1: Not 1; Site_2: Eco RI; Plasmid DNA
                    from the normalized library NCI_CGAP_G04 was prepared, and
                    ss circles were made in vitro. Following HAP purification,
                    this DNA was used as tracer in a subtractive hybridization
                    reaction. The driver was PCR-amplified cDNAs from a pool
                    of 5,000 clones made from the same library (clonings
                    1257096-1258631, 1459054-1470983, and 1475592-1476743).
                    Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      151 a 211 c 158 g 141 t
ORIGIN
1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyArg 17
|||||
466 TTATTTTAAATAAAATAAGGCTTTTATGCTCTGGGGAAGAGGGG 515
|||||
17 qAspArgCysLeuHisGluSerLysGlyArg 27
|||||
516 GCACCGCTGCTTACATACAGTCCAAAGCGCGCT 546
|||||
seq_name: qb_est2:BG820537
seq_documentation_block:
LOCUS          BG820537
DEFINITION    602782784F1 NCL_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4933675
5', mRNA sequence.
ACCESSION     BG820479
VERSION       BG820479
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE     1 (bases 1 to 788)
AUTHORS       NIH-MGC http://mgsa.ncbi.nih.gov/

```

```

5', mRNA sequence.
ACCESSION      BG820537
VERSION        BG820537.1 GI:14168124
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE      1 (bases 1 to 788)
AUTHORS        NIH-MGC http://mgsa.ncbi.nih.gov/
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished (1999)
COMMENT        Contact: Robert Strausberg, Ph.D.
                Email: cgapbs@mail.nih.gov
                Tissue Procurement: David N. Louis, M.D.
                cDNA Library Preparation: Life Technologies, Inc.
                DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LLAM10861 row: q column: 20
                High quality sequence stop: 783.
FEATURES        source
                  1..783
                  location/Qualifiers
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone IMAGE:4933675"
                    /clone_lib="NCL_CGAP_Brn67"
                    /tissue_type="anaplastic oligodendroglioma with ip/1q
                    loss"
                    /lab_host="pH108 (11 phage-resistant)"
                    /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: Not 1;
                    Site_2: SalI; Cloned unidirectionally. Primer: oligo dT.
                    Average insert size 2.3 kb. Constructed by Life
                    Technologies. Note: this is a NCL_CGAP library."
BASE COUNT     203 a 227 c 175 g 178 t
ORIGIN
1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyArg 17
|||||
435 TTCTACTTTAAATAAAATAAGGCTTTTATGCTCTGGGGAAGAGGGG 484
|||||
17 qAspArgCysLeuHisGluSerLysGlyArg 27
|||||
485 GGAAGCTTATGCTTAAATAAGTCCAAAGCGCGG 515
|||||
seq_name: qb_est2:BG820479
seq_documentation_block:
LOCUS          BG820479
DEFINITION    602782784F1 NCL_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4933675
5', mRNA sequence.
ACCESSION     BG820479
VERSION       BG820479
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE     1 (bases 1 to 788)
AUTHORS       NIH-MGC http://mgsa.ncbi.nih.gov/

```


TITLE JOURNAL COMMENT

National Institutes of Health, Mammalian Gene Collection (MGC)
unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
plate: L1AM10861 row: q column: 21
High quality sequence stop: 777.

FEATURES source

1..788
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:493676"
/clone_lib="MCL CGAP_Brn67"
/tissue_type="anaplastic oligodendroglioma with lp/19q
loss"
/lab_host="DH10B (11 phage-resistant)"
/note="Organ: brain; Vector: pMW-Sheep6; Site: 1, Not 1;
Site: 2, SalI. Cloned unidirectionally. Primer: oligo dI
Average insert size 2.3 kb. Constructed by Life
Technologies. Note: This is a MCL CGAP Library."
BASE COUNT 209 a 224 c 174 g 181 t
ORIGIN

alignment_scores:

Quality: 155.00 Length: 27
Ratio: 5.741 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

us-09-557-262-4_copy_246_272 x BG820479 ..

Align seq 1/1 to: BG820479 from: 1 to: 788

1 PhcTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyArg 17
|||||
435 TTCTACTTTAAACAAAGGGCTTTTACTGCTGGGAGACAAAGGGCGG 484

17 gAspArgCysLeuHisGluSerLysGlyArg 27
|||||
485 GGACCGCTGCTTACATGAGTCCAAAGGGCGG 515

seq_name: qb_est1.A1154059

documentation_block:

US A1154059 283 bp mRNA linear EST 30-Sep-1998
DEFINITION ud58f01.r1 Soares_NMPu Mus musculus cDNA clone IMAGE:1450105 5'
similar to TP 035310 035310 HEPAPAN SULFATE PRECURSOR 3'-UTRANSAMINYL
3'-O-SULFOTRANSFERASE-1 PRECURSOR .; mRNA sequence.

ACCESSION A1154059

VERSION A1154059.1 GI:3682528

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 283)
AUTHORS Marra, M., Hillier, L., Allen, M., Rowles, M., Dietrich, N., Eubouque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Slepecky, M., Tan, F., Underwood, K., Moore, R.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, P. and
Waterston, R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project
Washington University School of Medicine

4444 Forest Park Parkway, Box #501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:923421

Trace considered overall poor quality

Possible reversed clone: similarity on wrong strand

Seq primer: -28ml3 rev2 RT from Amersham

High quality sequence stop: 1.

FEATURES

source

1..283
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1450105"
/clone_lib="Soares_NMPu"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; 1st strand cDNA was prepared from
pregnant mouse uterus, and was then primed with a Not I -
oligo(dI) primer. Double stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Ronaldo."
BASE COUNT 80 a 30 c 64 g 59 t
ORIGIN

alignment_scores:

Quality: 152.00 Length: 27
Ratio: 5.630 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 95.296

alignment_block:

us-09-557-262-4_copy_246_272 x A1154059 ..

Align seq 1/1 to: A1154059 from: 1 to: 283

1 PhcTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyArg 17
|||||
98 TTCTACTTTAAACAAAGGGCTTTCTACTGCTGGGAGACAGTGGCAA 147

17 gAspArgCysLeuHisGluSerLysGlyArg 27
|||||

148 GGACCGCTGCTTACATGAGTCCAAAGGGCGG 178

seq_name: qb_est2.HB850108

seq_documentation_block:

US HB850108 291 bp mRNA linear EST 26-Sep-2000
DEFINITION ud58f01.r1 Soares mouse 28MS Mus musculus cDNA clone IMAGE:1416472
5' similar to TP 035310 035310 HEPAPAN SULFATE PRECURSOR .; mRNA sequence.

ACCESSION HB850108

VERSION HB850108.1 GI:10408447

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 291)
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

```

MGI:1092284
Seq primer: -40RP from Gibco
High quality sequence stop: 269.

FEATURES
SOURCE
1..291
/organism="Mus musculus"
/strain="G57BL/6J"
/db_xref="taxon:10090"
/cclone="IMAGE:3416472"
/cloned_lib="Soares mouse 3NDMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Vector: pUT10-pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I + oligo(dT) primer [5]. TGTTCAGATCGAAGTGAGACGGAGCCTTTTTTTTITTTTTTTT (Pharmac), digested with Not I and cloned into the Not I and Eco RI sites of the modified pU713 vector. RNA provided by Dr. Hertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      68 a    92 c    69 g    62 t
ORIGIN
alignment_scores:
Quality: 152.00          Length: 27
Ratio: 5.630             Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:
US-09-557-262-4_C24Y_246_272 x B3890529 ..
Align seq 1/1 to: BEF5010R from: 1 to: 291

1 PhcTyRheAsnlyshrlgysClphCyTcyLeuarjASbsctCYAr 17
|||||
192 TTCTACTTTAAACAACCAACCCCTTTCTACTGCTGCCGCACACGCGAA 241

17 qASpArcCysleullisgluserylslYArq 27
|||||
242 GCACCCTGCTTTACACGACACGCCAACAGCGCGC 272

seq_name: qb_est1.AA407647
seq_documentation_block:
LOCUS AA407647
DEFINITION EST009433 Mouse 7.5 dpc embryo ectoplacental cone cDNA library Mus musculus CDNA clone C0004A09 5' mRNA sequence.
ACCESSION AA407647.1 GI:2065848
VERSION AA407647
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus
KO.M.S.H., Threlkett,A., Horton,J.H., Wang,X., Gul,Y., Wang,X., Pryor,E., Paris,J., Wells-Smith,J., Fujiwara,H., Yotsumoto,S. and Nakashima,H.
Systematic analyses of mouse genes expressed in embryo implantation site
Unpublished (1997)
Other_FSTS: EST009432
Contact: Ko MSH
Center for Molecular Medicine and Genetics
Wayne State University
5047 Gullen Mall, Detroit, MI 48202
Tel: 3135776708
Fax: 3135776200
Email: mskomb@biosci.wayne.edu
Seq primer: M13 Reverse.
Location/Qualifiers
1..362
source

MCI:1092284
Seq primer: -40RP from Gibco
High quality sequence stop: 269.

FEATURES
SOURCE
1..291
/organism="Mus musculus"
/strain="G57BL/6J"
/db_xref="taxon:10090"
/cclone="IMAGE:3416472"
/cloned_lib="Soares mouse 3NDMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Vector: pUT10-pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I + oligo(dT) primer [5]. TGTTCAGATCGAAGTGAGACGGAGCCTTTTTTTTITTTTTTTT (Pharmac), digested with Not I and cloned into the Not I and Eco RI sites of the modified pU713 vector. RNA provided by Dr. Hertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      68 a    92 c    69 g    62 t
ORIGIN
alignment_scores:
Quality: 152.00          Length: 27
Ratio: 5.630             Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:
US-09-557-262-4_C24Y_246_272 x HRF5108 ..
Align seq 1/1 to: BER5010R from: 1 to: 291

1 PhcTyRheAsnlyshrlgysClphCyTcyLeuarjASbsctCYAr 17
|||||
192 TTCTACTTTAAACAACCAACCCCTTTCTACTGCTGCCGCACACGCGAA 241

17 qASpArcCysleullisgluserylslYArq 27
|||||
242 GCACCCTGCTTTACACGACACGCCAACAGCGCGC 272

seq_name: qb_est2.HC090529
seq_documentation_block:
LOCUS BC090529
DEFINITION mac07g02.y1 Soares mouse 3NDS Mus musculus cDNA clone IMAGE:3999170 5' similar to FR.G35310 HIFARAN SULFATE GLUCOSAMINYL 3-O-SULFOTRANSFERASE PRECURSOR /, mRNA sequence.
ACCESSION BC090529
VERSION BC090529.1 GI:12573092
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus
NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: rgs@fmail.nih.gov
This clone is available royalty-free through INEL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MCI:1496896
Seq primer: -40RP from Gibco
High quality sequence stop: 347.
Location/Qualifiers
1..362
source
```

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="ATCC (inhost):1363251"
/db_xref="taxon:10090"
/clone="r0004A09"
/clone_lib="Mouse 7.5 dpc embryo ectoplacental cone cDNA
library"
/sex="unknown"
/tissue_type="ectoplacental cone"
/dev_stage="embryonic day 7.5 postconception"
/lab_host="DH10B"
/notes="Organ: embryo, Vector: pSPOR11 (Life Technologies);
Site_1: SalI; Site_2: NotI; Total RNAs were extracted from
ectoplacental cone of 7.5-dpc embryos. The
double-stranded cDNA was synthesized from total RNAs with
an Oligo(dT) primer. The library was constructed by Minoru
S. H. Ko."
BASE COUNT      105 a  83 c   77 g   97 t
ORIGIN
alignment_scores:
    Quality: 152.00      Length: 27
    Ratio: 5.630        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 96.296
alignment_block:
    US-09-557-262-4_copy_246_272 x AA407647
Align seg 1/1 to: AA407647 from: 1 to: 362
1  PhcTyrPhcAsnLysThrGlyCysLeuArgAspSerGlyA: 17
|||||
12 TTCTACTTTTAAACAAACGAGGAGTTCCTATTTCTTCTGAGAAATAGTAA 61
17 gAspArgCysLeuHisGluSerLysGlyArg 27
|||||
62 CGACCCCTCTCTACACGAGTCCAAAGCCCG 92
seq_name: gb_est2.R0900452
seq_documentation_block:
LOCUS      R0900452               425 bp    mRNA    linear    EST 26-JAN-2001
DEFINITION mac06h02.Y1 Soares mouse 3NDMS Mus musculus cDNA clone
IMAGE:399195.5' similar to TF-03510 G3310 HEPARAN SULFATE
GLUCOSAMINYL 4-O-SULFOTRANSFERASE PPECHSOP ;, mRNA sequence.
ACCESSION  R0900452
VERSION    R0900452
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 425)
AUTHORS   NCI-CCAP http://www.nci.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: ccapbs-re@mail.nih.gov
            This clone is available royalty-free through LILM; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information
            MGI:1446023
Seq primer: -40RP from Gibco
High quality sequence stop: 378.
location/Qualifiers
    1..425
    /organism="Mus musculus"
    /strain="C57BL/6J"
    /db_xref="taxon:10090"
    /clone="IMAGE:399195"
    /clone_lib="Soares mouse 3NDMS"
    /sex="male"
FEATURES
    source

```

```

/tissue_type="Splicer"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Vector: pT7TD-Pac (Pharmacia) with a modified
polylinker; Site_1: NotI; Site_2: EcoRI; 2st strand cDNA
was primed with a NotI - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGGCGGCTGTGTGTGTGTGTGTGTGTGT
3']; double-stranded cDNA was ligated to EcoRI adaptors
(Pharmacia), digested with NotI and cloned into the NotI
and EcoRI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M. Fatima Honaldo."
BASE COUNT      104 a  124 c   93 g  104 t
ORIGIN
alignment_scores:
    Quality: 152.00      Length: 27
    Ratio: 5.630        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 96.296
alignment_block:
    US-09-557-262-4_copy_246_272 x R0900452
Align seg 1/1 to: R0900452 from: 1 to: 425
1  PhcTyrPhcAsnLysThrGlyCysLeuArgAspSerGlyA: 17
|||||
192 TTCTACTTTTAAACAAACGAGGCTTCCTATTTCTTCTGAGAAATAGTAA 241
17 gAspArgCysLeuHisGluSerLysGlyArg 27
|||||
242 CGACCCCTCTCTACACGAGTCCAAAGCCCG 272
seq_name: gb_est2.W62484
seq_documentation_block:
LOCUS      W62484               490 bp    mRNA    linear    EST 07-JUN-1996
DEFINITION md72c12.r1 Soares mouse embryo NbMEL3.5 14.5 Mus musculus cDNA
clone IMAGE:373942.5' similar to PIR:A49733 A49733
glycosaminoglycan N-acetylglucosaminyl N-deacetylase ;, mRNA
sequence.
ACCESSION  W62484
VERSION    W62484.1   GI:1369243
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 490)
AUTHORS   Maria M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,
            Geisel S., Kuraba T., Lary M., Le M., Martin J., Morris M.,
            Schellenberg K., Steptoe M., Tan F., Underwood K., Moore H.,
            Theising B., Wylie T., Kennon G., Soares B., Wilson R. and
            Waterston R.
TITLE     The WashU-HMI Mouse EST Project
JOURNAL    Unpublished (1996)
COMMENT    Contact: Marra M/Mouse EST Project
            WashU-HMI Mouse EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LILM; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:235374
Seq primer: ETPrimer
High quality sequence stop: 344.
location/Qualifiers
    1..490
    /organism="Mus musculus"
FEATURES
    source

```


WashU-HIMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.wustl.edu

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info:image.llnl.gov) for further information.

MCL:286012

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 490.

FEATURES
Source

Location/Qualifiers

1..508

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:475268"

/sex="unknown"

/issue_type="embryo"

/dev_stage="13.5, 14.5dpc total fetus"

/lab_host="DH10B"

/note="Vector: p7130-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA
was primed with a Not 1 - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGGCGGCGCTTTTITTTTTTTTTT
T 3']; on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minora Ko, Wayne
State Univ., from 2 14.5dpc embryos) cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not 1 and
cloned into the Not 1 and Eco RI sites of the modified
p773 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M. Fatima Ronaldo."

BASE COUNT 136 a 134 c 114 g 124 t

ORIGIN

alignment_scores:

Quality: 152.00 Length: 27
Ratio: 5.630 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:

US-09-557-262-4_copy_246_272 x AAC41985

ign seq 1/1 to: AAC41885 from: 1 to: 568

1 PheTyPheAsnLysThrLysGlyPheTyPheCysLeuArgAspSerGlyAr 17

|||||
196 TTC TAC TTT AAC AAAA CCAAGGCGCTTCTACTGCTGCGGACAGTGGCAA 245

17 qAspArgCysLeuHisGluSerLysGlyArg 27

|||||
246 GGACCGCTGCTTACACAGATCCAAAGGCGG 276

seq_name: qb_est2:BE851982

seq_documentation_block:

LOCUS BE851982 555 bp mpna linear EST 26-SEP-2000
DEFINITION w12e07.x1 Soares mouse mDMS Mus musculus cDNA clone IMAGE:3416484
3' similar to TR:035310.035310 HEPARAN SULFATE GLUCOSAMINYL
3'-O-SULFOTRANSFERASE PRECURSOR ?, mRNA sequence.

ACCESSION BE851982

VERSION BE851982.1 GI:10310321

KEYWORDS EST

SOURCE house mouse

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.

1 (bases 1 to 555)

AUTHORS NC1-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL

COMMENT

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

This clone is available royalty free through LLNL; contact the

IMAGE Consortium (info:image.llnl.gov) for further information.

MGI:1092296

High quality sequence stop: 474.

FEATURES

Location/Qualifiers

1..555

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="IMAGE:3416484"

/sex="male"

/issue_type="Spleen"

/dev_stage="4 weeks"

/lab_host="DH10B"

/note="Vector: p7130-Pac (Pharmacia) with a modified
polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA
was primed with a Not 1 - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGGCGGCGCTTTTITTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not 1 and cloned into the Not 1
and Eco RI sites of the modified p773 vector. RNA
provided by Dr. Norihiro Jordan library went through
three rounds of normalization, and was constructed by
Bento Soares and M. Fatima Ronaldo."

BASE COUNT 173 a 111 c 114 g 157 t

ORIGIN

alignment_scores:

Quality: 152.00 Length: 27
Ratio: 5.630 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.296

alignment_block:

US-09-557-262-4_copy_246_272 x BE851982/rev

Align seq 1/1 to reverse of: BE851982 from: 1 to: 555

1 PheTyPheAsnLysThrLysGlyPheTyPheCysLeuArgAspSerGlyAr 17

|||||

486 TTC TAC TTT AAC AAAA CCAAGGCGCTTCTACTGCTGCGGACAGTGGCAA 437

17 qAspArgCysLeuHisGluSerLysGlyArg 27

|||||

436 GGACCGCTGCTTACACAGATCCAAAGGCGG 406



GM 6 05-04-557 262-2_copy_250_276.p2n.rst * out_format : fts
 Date: Jul 17, 2002 3:40 AM
 About: Results were produced by the GenCore software, version 4.5.
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frames-p2n.model -DEV=xlp
 -O=/cgn2.1/USPTO_spool/US06557262/runat_16070002_161204_2554/app_query fasta.1 1566
 -DB-EST -QFMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LAPPEXT=0.000 -LQPEXT=0.000 -QCAPOP=4.500
 -QCAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FCAPOP=6.000
 -FCAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEX=7.000 -START=1 -MATRIX=bloms62 -TRANS=bloms62.caf
 -LIST=45 -DOALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -TUR_MIN=0
 -ALIGN=15 -MODE=LOCAL -DIFFMT=plis -NORM=ext -HEAPSIZ=500
 -MINLEN=0 -MAXLEN=20000000 -USER=US0557262_accn1.1.7321
 -NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMECUT=120 -WARN_TIMECUT=40
 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-557-262-2_Copy_250_276
 Query length: 27
 Database: EST:
 Database sequences: 13736207
 Database length: -1841457050
 Search time (sec): 5635 550000

score_list:

Sequence	Strd	Orig	ZScore	Escore	len	Documentation
gb_est1: A1154059	+	155.00	419.41	4.1e-14	283	! A1154059 ud58101 r1 Soares_NMPu
gb_est2: B8501068	+	155.00	419.18	4.3e-14	291	! B8501068 uw12601 y1 Soares_mous
gb_est2: B0290524	+	155.00	417.45	5.3e-14	454	! B0290524 mac02902 y1 Soares_mou
gb_est1: A4407647	+	155.00	417.38	5.3e-14	362	! A4407647 EST009333 Mouse 7.5 dpc
gb_est2: B0290452	+	155.00	416.05	6.3e-14	425	! B0290452 mac06h02 y1 Soares_mou
gb_est2: B622484	+	155.00	414.87	7.3e-14	400	! B622484 md72-12 r1 Soares_mous
gb_est2: B0122797	-	155.00	414.84	7.4e-14	402	! B0122797 L053506 3 NIA Mouse N
gb_est2: B0124136	-	155.00	414.64	7.5e-14	504	! B0124136 L053506 3 NIA Mouse N
gb_est1: A4407566	+	155.00	414.59	7.6e-14	507	! A4407566 EST009333 Mouse 7.5 dpc
gb_est1: A4404185	+	155.00	414.57	7.6e-14	508	! A4404185 m305011 r1 Soares_mous
gb_est2: B851982	-	155.00	414.84	8.4e-14	555	! B851982 uw12607 x1 Soares_mous
gb_est1: A441469	+	155.00	413.45	8.8e-14	582	! A441469 EST291519 Normalized B
gb_est1: A538931	+	155.00	412.85	9.5e-14	626	! A538931 col06c03 NIA Mouse B
gb_est2: B012413	+	155.00	412.60	9.8e-14	645	! B012413 EST272558 Normalized B
gb_est2: B0123143	+	155.00	412.57	9.8e-14	647	! B0123143 L0520808 3 NIA Mouse N
gb_est1: B8554892	+	155.00	412.38	1.0e-13	662	! B8554892 B8554892 RIKEN full-le
gb_est1: A1013394	+	155.00	412.28	1.0e-13	670	! A1013394 EST208069 Normalized B
gb_est2: B029238	-	155.00	412.17	1.0e-13	679	! B029238 mac06h02 x1 Soares_mou
gb_est1: A1266939	+	155.00	411.42	1.1e-13	744	! A1266939 qj08h03 x1 Sugano_mou
gb_est1: A1971626	+	152.00	404.58	2.7e-13	626	! A1971626 AV694733 GRC Homo sapi
gb_est2: B02820537	+	152.00	404.26	2.9e-13	651	! B02820537 w06611 x1 NCI_CGAP_H
gb_est2: B02820537	+	152.00	402.73	3.5e-13	783	! B02820537 602782876F1 NCI_CGAP_H
gb_est2: B02820479	+	152.00	402.68	3.5e-13	788	! B02820479 602782876F1 NCI_CGAP_H
gb_est1: A1971626	+	149.00	392.51	1.3e-12	992	! A1971626 AV236060 Normalized B
gb_est1: A1971626	+	148.00	392.20	7.1e-13	403	! A1971626 EST291517 Normalized B
gb_est2: B0428497	+	148.00	395.70	8.1e-13	483	! B0428497 qd410604 x1 NIH_HD_Xcr
gb_est2: B038480	+	148.00	395.00	9.4e-13	526	! B038480 qd410604 x1 Wellcome_C
gb_est2: B038674	+	148.00	394.51	1.0e-12	558	! B038674 qd410604 x1 Wellcome_C
gb_est1: A1971626	+	148.00	394.28	1.0e-12	574	! A1971626 d411a06 x1 Xenopus lae
gb_est1: A1971626	+	148.00	393.79	1.1e-12	609	! A1971626 w176101 x1 NCI_CGAP_H
gb_est2: B0428497	+	146.00	387.74	2.4e-12	451	! B0428497 qd410604 x1 NCI_CGAP_H
gb_est1: A1971626	+	139.00	373.96	1.4e-11	334	! A1971626 qd410604 x1 NCI_CGAP_H
gb_est2: B0676274	+	139.00	368.83	3.2e-11	621	! B0676274 qd410604 x1 NCI_CGAP_H
gb_est2: B0676274	+	137.00	367.49	3.2e-11	375	! B0676274 yb40007 r1 Stratagene Fe
gb_est1: A1431882	+	135.00	358.52	1.0e-10	570	! A1431882 t126609 x1 NCI_CGAP_H
gb_est2: B0348514	+	132.00	353.27	7.0e-10	496	! B0348514 d494102 y1 Wellcome_C
gb_est2: B0348514	+	132.00	348.89	3.5e-10	672	! B0348514 B0325717 MF0158A cDNA
gb_est2: B0325717	-	132.00	340.71	9.9e-10	928	! B0325717 Tetradodon nigroviridis
gb_est2: B0325717	-	128.00	340.12	1.1e-09	512	! B0325717 H400010023 NIA Mouse B
gb_est2: B0325717	+	128.00	339.15	1.2e-09	576	! B0325717 qd50604 x1 NCI_CGAP_H

gb_est1: A1582332 + 128.00 339.56 1.4e-09 + A1582332 t377501 x1 NCI_CGAP
 gb_est1: A1971645 + 128.00 339.51 1.4e-09 + A1971645 w06611 x1 NCI_CGAP
 gb_est2: B0603066 + 128.00 337.57 1.5e-09 + B0603066 H3901r12 3 NIA Mous
 gb_est2: B14307496 + 125.00 328.56 4.5e-09 + B14307496 c2341872411 NCI_CGA
 gb_est1: B8786419 + 126.00 319.45 1.5e-09 + B8786419 B8786419 RIKEN full

seq_name: gb_est1: A1154059
 seq_documentation_block:
 Locus A1154059 chr11:140000000-140000000 EST 30-SEP-1998
 DEFINITION ud58101 r1 Soares_NMPu Mus musculus cDNA clone IMAGE1450105 5'
 Similar to TR035310 035310 HEPARAN SULFATE D-GLUCOSAMINYL
 3-O-SULFOTRANSFERASE-1 PRECURSOR 1, mRNA sequence.
 ACCESSION A1154059
 VERSION A1154059.1 GI:3682538
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Scurionathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 283)
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubouque, J.,
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Mariani, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, R.,
 Theisinger, B., Wyllie, J., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT WashU-HMI Mouse EST Project
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mgeese@wustl.wustl.edu
 This clone is available publicly free through the IMAGe Consortium (info@image.llnl.gov) for further information.
 IMAGe Consortium (info@image.llnl.gov) for further information.
 MGI:923421

Trace considered overall poor quality
 Possible reversed strand similarity on w06611 strand
 Seq primer: 28m13 rev2 ET from Amersham
 High quality sequence stop: 1.
 FEATURES
 location/Qualifiers
 .. 283
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE1450105"
 /clone_lib="Soares_NMPu"
 /sex="female"
 /dev_stage="adult"
 /lab_host="RH10B"
 /note="organ. uterus; Vector: p773D-Pac (Pharmacia) with
 a modified polylinker; 1st strand cDNA was prepared from
 a pregnant mouse uterus, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified p773
 vector. Library is normalized. Library was constructed by
 Renato Soares and M. Fatima Bonaldi."

RASP COUNT 80 4 80 c 64 g 59 t
 ORIGIN
 alignment_scores:
 Quality: 155.00 Length: 27
 Ratio: 5.741 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 HS 09 557 262-2_Copy_250_276 x A1154059
 A1154059 1/1 to: A1154059 from: 1 to: 283


```

sequence.
ACCESSION W62484
VERSION W62484.1 GI:1369243
KEYWORDS EST
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Pukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 490)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, J., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty free through HMI; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:235374
Seq primer: EST primer
High quality sequence stop: 344.
FEATURES
Location/Qualifiers
1..490
/organism="Mus musculus"
/strain="C57Bl/6J"
/db_xref="taxon:10090"
/clone_image="373942"
/clone_lib="Soares mouse embryo NHEK13 5 14 5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/notes="Vector: pT7T3-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer 15',
TGTTCACCAATCGAAGTGGAGCGCCGCCGGAATTTTCTTTTCTTTTCTTTT
1 3', on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Micoira KO, Wayne
State Univ, from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M. Fatima Bonaldo."
BASE COUNT 127 a 129 c 112 g 122 t
ORIGIN

alignment_scores:
Quality: 155.00 Length: 27
Ratio: 5.741 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-557-262-2_COPY_250_276 x W62484 1 to 490
Align seq 1/1 to W62484 from 1 to 490

1 PheTyrPheAsnLysThrLysGlyPheTyrCysLeuArgAspSerGlyly 17
|||||
196 TTCTACTTTTAAACAAACCAAGCGCTTCTACTTGGCTGGCGACAGTGGCAA 245

17 SASPArcCysLeuLisGluSerLysGlyArc 27
|||||
246 GGACCGCTGCTTACACAGATCCAAAGCGCG 276

```

```

seq_name: qb_est2:BM122797
seq_documentation_block:
LOCUS BM122797 492 bp mRNA linear EST OF FEB 2002
DEFINITION L0515C10-3 NIA Mouse Newborn Heart cDNA Library Mus musculus cDNA
clone L0515C10 3', mRNA sequence.
ACCESSION BM122797
VERSION BM122797.1 GI:17106565
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Pukaryota; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 492)
Piao, Y., Karpiuk, G.J., Dudekula, D.B., Qian, Y., Pantano, S., Lim, M.K.
and Ko, M.S.H.
TITLE Systematic Analyses of NIA Mouse Newborn Heart cDNA Library
JOURNAL Unpublished (2001)
COMMENT Contact: Howard R. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@niasun.qrc.nia.nih.gov
Plate: 0515 row C column: 10
Seq primer: -21M13 Forward
High quality sequence stop: 492
POLYA=yes.
FEATURES
Location/Qualifiers
1..492
/organism="Mus musculus"
/strain="C57Bl/6J"
/db_xref="niaEST:L0515C10-3"
/db_xref="taxon:10090"
/clone_image="L0515C10"
/clone_lib="NIA Mouse Newborn Heart cDNA Library"
/tissue_type="Newborn Heart"
/dev_stage="DH10B"
/lab_host="DH10B"
/notes="Organ: heart; Vector: pSPORT1 (Invitrogen); Site 1:
SalI; Site 2: NotI; Mouse cDNA project by the Laboratory
of Genetics, National Institute on Aging (NIA), Intramural
Research Program, NIH (http://niasun.qrc.nia.nih.gov/CDNA);
double-stranded cDNAs were synthesized with an oligo(dT)
primer [Invitrogen: 5'-
pCACTAGTCTAGATGCGAGCGGCGGCTTTTCTTTTCTTTT 3'] from
24.9 microgram of total RNA, treated with T4 DNA
polymerase and purified by ethanol-precipitation. The
cDNAs were ligated to lone-linker LL-Sal3 (ref.
development 127:1737-1749 (2000) [PMID: 10725249]),
purified by phenol/chloroform, and separated from free
linkers by centrifugation 100. Then, the cDNAs were digested
with SalI and NotI enzymes, and cloned into SalI and NotI
site of pSPORT1 plasmid vector. The DH10B E. coli host was
transformed with ligation mixture by the chemical method.
The average insert size is about 1.8 kb. The library was
constructed by Yulan Piao (NIA)."
BASE COUNT 157 a 102 c 91 g 142 t
ORIGIN

alignment_scores:
Quality: 155.00 Length: 27
Ratio: 5.741 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-557-262-2_COPY_250_276 x BM122797 rev
Align seq 1/1 to reverse of: BM122797 from 1 to 492

```

17 sAspArqCysLeuHisGluSerLysGlyArg 27
 |||||
 438 GGACCGCTGCTTACAGAGTCCAAAGCGCG 408

seq_name: qb_est2:BM124136

seq_documentation_block:

LOCUS BM124136 504 bp mRNA linear EST 01-FEB-2002
 DEFINITION L0535G06-3 NIA Mouse Newborn Heart cDNA library Mus musculus cDNA
 clone L0535G06 3', mRNA sequence.

ACCESSION BM124136

VERSION BM124136 1 GI:17107904

KEYWORDS EST

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

1 (bases 1 to 504)

Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Pantano, S., Lim, M.K.

and K.O.M.S.H.

Systematic Analyses of NIA Mouse Newborn Heart cDNA Library

UNPUBLISHED (2001)

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224 6830, USA

Email: cdnaelgusun.grc.nia.nih.gov

Plate: L0535 row: C column: 06

Seq primer: -21M13 Forward

High quality sequence stop: 504

POLYA=yes.

FEATURES

source

1..504

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="niaESP:L0535G06-3"

/db_xref="taxon:10090"

/clone="L0535G06"

/tissue_type="Newborn Heart"

/dev_stage="Newborn"

/lab_host="DH10B"

/note="Organ heart, Vector pSPORT1 (Invitrogen); Site_1:

SalI; Site_2: NotI; Mouse cDNA project by the laboratory

of Genetics, National Institute on Aging (NIA), Intramural

Research Program, NIH (<http://lgsun.grc.nia.nih.gov/cDNA>).

Double-stranded cDNAs were synthesized with an oligo(dT)

primer (Invitrogen): 5'-

PGACTAGTCTAGATGGAGCGCGCGCTTTTCTTTT-3'}

from

24.9 microgram of total RNA, treated with T4 DNA

polymerase, and purified by ethanol-precipitation. The

cDNAs were ligated to lone-linker L1-Sal3 (Ref.

Development 127 1747-1749 (2000) [PMID: 10725240]).

Purified by phenol/chloroform, and separated from free

linkers by (centrifuge 100 Then, the cDNAs were digested

with SalI and NotI enzymes, and cloned into SalI and NotI

site of pSPORT1 plasmid vector. The DH10B E. coli host was

transformed with ligation mixture by the chemical method.

The average insert size is about 1.8 kb The library was

constructed by Yulan Piao (NIA)."

BASE COUNT

ORIGIN

159 a 104 c 96 g 145 t

alignment_scores:

Quality: 155.00 Length: 27

Ratio: 5.741 Gaps: 0

Percent Similarity: 100.000 Percent identity: 100.000

alignment_block:

US-09-557-262-2_copy_250_276 x BM124136/rev

Align seq 1/1 to reverse of: BM124136 from: 1 to: 504

1 PheTyPheAsnLysThrLysGlyPheTyPheCysLeuAqAspSerGly 17
 |||||
 488 TTCTACTTTAACAAACCAACGGCTTCTACCTGCGGACAGTGGCA 439

17 sAspArqCysLeuHisGluSerLysGlyArg 27

|||||

438 GCACCGCTGCTTACAGAGTCCAAAGCGCG 408

seq_name: qb_est1:AA407566

seq_documentation_block:

LOCUS AA407566

DEFINITION EST000431 Mouse 7.5 dpc embryo ectoplacental cone cDNA library Mus
 musculus cDNA clone C0004A08 5', mRNA sequence.

ACCESSION AA407566

VERSION AA407566 1 GI:2055846

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

1 (bases 1 to 507)

Ko, M.S.H., Threalt, T.A., Horton, J.H., Wang, X., Wang, X., Pryor

, E., Paris, J., Wells-Smith, J., Fujiwara, H., Yotsumoto, S. and

Nakashima, H.

Systematic analyses of mouse genes expressed in embryo implantation

site

UNPUBLISHED (1997)

Contact: Ko MSH

Center for Molecular Medicine and Genetics

Wayne State University

5047 Gullen Mall, Detroit, MI 48202

Tel: 3135776708

Fax: 3135776200

Email: mskocmb.biosci.wayne.edu

Seq primer: M13 Reverse.

location/Qualifiers

1..507

source

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="ATCC (inhost):1553226"

/db_xref="taxon:10090"

/clone="C0004A08"

/clone_lib="Mouse 7.5 dpc embryo ectoplacental cone cDNA

library"

/sex="unknown"

/tissue_type="ectoplacental cone"

/dev_stage="embryonic day 7.5 postconception"

/lab_host="DH10B"

/note="Organ: embryo; Vector: pSPORT1 (Life Technologies);

Site_1: SalI, Site_2: NotI. Total RNAs were extracted from

ectoplacental cone of 7.5-dpc embryos. The

double-stranded cDNA was synthesized from total RNAs with

an oligo(dT) primer. The library was constructed by Minoru

S. H. Ko."

BASE COUNT 145 a 98 c 106 g 156 t 2 others

ORIGIN

alignment_scores:

Quality: 155.00 Length: 27

Ratio: 5.741 Gaps: 0

Percent Similarity: 100.000 Percent identity: 100.000

alignment_block:

US-09-557-262-2_copy_250_276 x AA407566

align seq 1/1 to: AA407566 from: 1 to: 507

1 PheTyPheAsnLysThrLysGlyPheTyPheCysLeuAqAspSerGly 17

12 TCTACATTAAACAAAGAGTGTATGCTAGAGAAATGACMA 61
 |||

17 SASPARgCysLeuHisGluSerIysGlyArg 27
 |||

62 GCACGCTGTCTTACAGAGTCAAAAGGCTGG 92
 |||

seq_name: qb_est1:AA041885

seq_documentation_block:

LOCUS AA041885 508 bp mRNA linear EST 03-SEP-1996
 DEFINITION mJ05g11.r1 Soares mouse embryo NM013.5 14.5 Mus musculus cDNA
 clone IMAGE:475268 5' similar to PIP A49733 A49734
 glycosaminoglycan N acetylglucosaminyl N deacetylase , mRNA
 sequence.

ACCESSION AA041885

VERSION AA041885.1 GI:1520040

WORDS EST

ORIGIN house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 508)
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Stepien, R., Tan, F., Underwood, K., Moore, B.,
 Theising, H., Wylic, T., Lennon, G., Soares, R., Wilson, R. and
 Waterston, K.

TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.wustl.edu

This clone is available royalty-free through LLNL : contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MC1:286312

Seq primer: 28M13 rev2 from Amersham

High quality sequence stop: 490.

FEATURES

source

1..508
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:475268"
 /clone_lib="Soares mouse embryo NM013.5 14.5"
 /sex="unknown"
 /tissue_type="embryo"
 /dev_stage="13.5-14.5dpc total fetus"
 /lab_host="DH10H"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not 1; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not 1 oligo(dT) primer [5',
 TGTTCACCATCGAGTCGAGCGCCGCCGAAATTTTTTTTTTTTTTTTTTT
 T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
 State Univ., from 2 1; double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not 1 and
 cloned into the Not 1 and Eco RI sites of the modified
 pT7T3 vector. Library went through one round of
 normalization, and was constructed by Bento Soares and
 M.Fatima Bonaldo."

BASE COUNT 136 a 134 c 114 q 124 t

ORIGIN

alignment_scores:

Quality: 155.00 Length: 27
 Ratio: 5.741 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-557-262-2_copy_250_276 x AA041885

Align seg 1/1 to: AA041885 from: 1 to: 508

1 PheTyrPheAsnLysThrIysGlyPheTyrCysLeuAraAspSerGlyArg 17
 |||

196 TTCACATTAAACAAAGAGTGTATGCTAGAGAAATGACMA 245
 |||

17 SASPARgCysLeuHisGluSerIysGlyArg 27

|||||

246 GCAAGCTGTCTTACAGAGTCAAAAGGCTGG 276
 |||

seq_name: qb_est2:BE851982

seq_documentation_block:

LOCUS BE851982 555 bp mRNA linear EST 26-SEP-2000
 DEFINITION uwl2e07.x1 Soares mouse 3BMS Mus musculus cDNA clone IMAGE:3416484
 3' similar to TR-045410 045410 HEPATAN SULFATE GLUCOSAMINYL
 3-O-SULFOHANSFERASE PRECURSOR ; mRNA sequence.

ACCESSION BE851982

VERSION BE851982.1 GI:10310321

WORDS EST

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 555)
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC-I-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

TITLE Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Other ESTs: uwl2e07.y1

Contact: Robert Strausberg, Ph.D.

Email: cqa@bbs.fda.nih.gov

This clone is available royalty-free through LLNL : contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MC1:1092296

High quality sequence stop: 474.

FEATURES

source

1..555
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:3416484"
 /clone_lib="Soares mouse 3BMS"
 /sex="male"
 /tissue_type="Spleen"
 /dev_stage="4 weeks"
 /lab_host="DH10H"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not 1; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not 1 oligo(dT) primer [5',
 TGTTCACCATCGAGTCGAGCGCCGCCGAAATTTTTTTTTTTTTTTTTTT
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not 1 and cloned into the Not 1
 and Eco RI sites of the modified pT7T3 vector. RNA
 provided by Dr. Bertrand Jordan. Library went through
 three rounds of normalization, and was constructed by
 Bento Soares and M.Fatima Bonaldo."

BASE COUNT 173 a 111 c 114 q 157 t

ORIGIN

alignment_scores:

Quality: 155.00 Length: 27
 Ratio: 5.741 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-557-262-2_copy_250_276 x BE851982/rev

17 sAspArqCysLeuHisGluSerLysGlyArg 27
|||||
438 GGACCGCTGCTTACACGACTCCAAAGGCCCG 408



11 GCATACAGCTAATAAAGAGGAGGACCACTTATACAGGACAGACCACTTACAG 60

81 rclncllyLeuclyl-TrpTyrLeuThrGlnMet_ProPheSerSerProHle 97

61 CCAATAGCGCTGCTTGGTAACTTAAAGATGAGGCTTCTCCCTGAGCTCA 110

97 scGlnLeuThrValGlnLysThrProAlaTyrPheThrSerProLysValP 114

111 ACAATATATGAGAGAGAGAAAGAGGATATTATATATATATATATATATAT 160

114 roGluArqGlnIleHisSerMetAsnProThrIleArqLeuLeuLeuLeu 130

161 CTGAGAGAACTCAGACAGTCAAGGCTCAACCTCACTGCTGCTGCTTATCCTC 210

131 ArqAspProSerGluArqValLeuSerAspTyrThrGlnValLeuTyrAs 147

211 AATAATATATAG 260

147 nHisLeuGlnLysHisLysProTyrProPheThrGlnValLeuLeuLeu 164

261 CCAATATATATAG 310

164 rAspGlyArqLeuAsnLeuAspTyrLysAlaLeuAsnArqSerLeuTyr 180

311 GCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 460

181 HisAlaHisMetLeuAsnTrpLeuArqPhePheProLeuGlyHisIleH 197

361 CATGCACACAT 410

197 sIleValAspGlyAspArgLeuIleArqAspProPheProGluIleGlnI 214

411 CATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 460

214 ysValGluArqPheLeuLysLeuSerProGlnIleAsnAlaSerAsnPhe 230

461 ACGTCGAAGAGATCTCTCAAGCTTCTCCAGATCAACGCTCCCAACTTC 510

231 TyrPheAsnLysThrLysGlyPheTyrGlyLeuArqAspSerGlyLysAs 247

511 TACTTTAACAATAAT 560

247 ParGlySerHisGlySerLysGlyArqAlaHisProGlnValAspProI 264

561 CAGTGTAT 610

264 ysLeuLeuAspLysLeuIleIleLysLysLysLysLysLysLysLysLys 280

611 AACTTACTGATGAATAACTGCAAGCAATCTTCTCACTGCAAGCAATGAAGAATTT 660

281 PheLysLeuValGlyArqThr 287

661 TTAAAT 681

seq_name: gb_est1:A1971626

seq_documentation_block:

LOCUS A1971626 651 bp mRNA linear EST 08-MAR-2000
 DEFINITION wr06c11 x1 NCI_CGAP_c66 Homo sapiens cDNA clone IMAGE:2480756 3' similar to TR-014792 G34792 HEPARAN SULFATE 3 O-SULFO1TRANSPIRASE-1 PRECURSOR. ; mRNA sequence.

ACCESSION A1971626

VERSION A1971626.1 GI:5768452

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Fukuyama, Metazoa, Chordata; Crustacea, Vertebrata, Euteleostomi,

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 651)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image.html
 Insert length: 1020 Std Error: 0.00
 Seq primer: 400P from Gibco

High quality sequence stop: 456.
 Location/Qualifiers

FEATURES

source

1..651
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2480756"
 /clone_lib="NCI_CGAP_C66"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; plasmid DNA from the normalized library NCI_CGAP_C66 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 151 a 211 c 158 g 131 t
 CPTGIN

alignment_scores:

Quality: 1045.00 Length: 214
 Ratio: 5.098 Gaps: 0
 Percent Similarity: 96.244 Percent Identity: 89.202

alignment_block:

US-09-557-262-2_COPY_21_307 x A1971626

Align seg 1/1 to: A1971626 from: 1 to: 651

75 AsPTrpGluGlnHisTyrSerGlnGlyLeuGlyTyrTyrLeuThrGlnMe 91
 1 GACTGGGAGGAGGCTTACAGCGAGCGCTTGGGCTGGTACCTCAGCCAGAI 50
 91 tProPheSerSerProHisGlnLeuThrValGlnLysThrProAlaTyrP 108
 51 GCGCTTCT 100
 108 heThrSerProLysValProGluArqGlnIleHisSerMetAsnProThrIle 124
 101 TCACGTCGCCCAAGAGTCCCTGACCGAGCTCTACAGCAAGAACCCGCTCCATC 150
 125 ArqLeuLeuLeuLeuLeuArqAspProSerGluArqValLeuSerAspTy 141
 151 CCGCTGCT 200
 141 rThrGlnValLeuTyrAsnHisLeuGlnLysHisLysLysProTyrProI 158
 201 CACCCAAAGTGTATTATATATATATATATATATATATATATATATATATAT 250
 158 LeGluAspLeuLeuLeuMetArqAspGlyValGlnAsnLeuAspTyrLysAla 174
 251 TCGAGGAGATTCCTCGGCGCGGAGCGGAGGCTTAATCTCTCTCTCTCTCTCT 300
 175 LeuAsnArqSerLeuTyrHisAlaHisMetLeuAsnTrpLeuArqPhePh 191
 301 CTCAACTGCAATCT 350

191 eProLeuGlyHisIleHisIleValAspGlyAspArgLeuValArgAsp 208
 351 CCGGTGGCCACATCCACATTTGGAGCGGACCGCTCATAGGAGCC 400
 208 rPheProGluIleGlnIysValGluArgPheLeuLysLeuSerProGln 224
 401 CCTTCCTCGATCCAAAGGTCAGAGGTTCCCTAAAGCTGTCGGCCAC 450
 225 IleAsnAlaSerAsnPheTyrPheAsnLysThrLysGlyPheTyrCysLe 241
 451 ATCAATGCTCGAATCTCTATCTTAAACAACAACCAAGGCTTTTACGGCT 500
 241 uArgAspSerGlyLysAspArgCysLeuHisGluSerLysGlyArgAlaH 258
 501 GCGGACAGCGCGCGGAGCGTGTGTACATAGTCCAAAAGAGCGTGGCC 550
 258 isProGluValAspProLysLeuAspLysLeuHisGluTyrPheHis 274
 551 ACCCCCAAGTCATCCCAAACTACICATAAATACGACGAATAGTTTCAT 600
 275 GluProAsnLysLysPhePheLysLeuValGlyArgThr 287
 601 GAGCCATAAAGATGTTCTCGAGCTTGTCGAGCAACA 639

seq_name: gb_est2:BF940710

seq_documentation_block:

LOCUS BF940710 651 bp mRNA linear EST 30-MAR-2001
 DEFINITION nae24b08.x1 NCI_CGAP_Ov18 Homo sapiens cDNA clone IMAGE:3435999 3' similar to TP-014792 G14792 HEPARAN SULFATE 3-O SULFOTRANSFERASE-1 PRECURSOR. ;, mRNA sequence.
 ACCESSION BF940710
 VERSION BF940710.1 GI:12358030
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 651)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: ccapbs-femail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
 Seq primer: -400p from Gibco
 High quality sequence stop: 470.
 FEATURES
 Location/Qualifiers
 1..651
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3435999"
 /clone_lib="NCI_CGAP_Ov18"
 /tissue_type="fibrotheoma"
 /lab_host="D1010B (phage-resistant)"
 /note="Organ: ovary; Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; 1st strand cDNA was primed with Not 1 - oligo(dT) primer [5' TGTTACCAATCTCAAGTCGAGCGCGCGCCACACATTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not 1 and cloned into the Not 1 and Eco RI sites of the modified p773 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Honaldo."

BASE COUNT

149 a 210 c 163 g 129 t

ORIGIN

alignment_scores:
 Quality: 1028.00 Length: 213
 Ratio: 5.064 Gaps: 0
 Percent Similarity: 95.405 Percent Identity: 86.854
 alignment_block:
 US-09-557-262-2_copy_21_307 x BF940710
 Align seq 1/1 to: BF940710 from: 1 to: 651
 75 AsPTrpGluGluHisIlySerGlnGlyLeuGlyTyrPheLeuThrGlnMe 91
 1 GACTGGAGAGAGATTAAGAAAGATTTGGCTTGGCTTGGCTTGGCTTGGCT 50
 91 tPhePheSerSerProHisGlnLeuThrValGluLysThrProAlaLysP 108
 51 GCGCTTCCTCGGCGCCACACACCTCCACAGTGGAGAACACCCCGCGGAT 100
 108 heThrSerProLysValProGluArgIleHisSerMetAsnProThrIle 124
 101 TCAGCTGGGCTAAAGTGGCTGGAGGAGCTTACACAGCAAGACCGGCTCA 150
 125 ArgLeuLeuLeuIleLeuArgAspProSerGluArgValLeuSerAsp 141
 151 GCGTGTGTATGTTATGTTATGTTATGTTATGTTATGTTATGTTATGTTAT 200
 141 rThrGlnValLeuTyrAsnHisLeuGlnLysHisLysPheTyrProPro 158
 201 CACCAAGCTGTTCTACACACCATGCGACAGCAACACCCCTACCCGCCA 250
 158 LeGluAspLeuLeuMetArgAspGlyArgLeuAsnLeuAspTyrLysAla 174
 251 TCGAGGAGTCTGCTGGCGGCGACGCGCTCAATGCGACTACAAAGGCC 300
 175 LeuAsnArgSerLeuTyrHisAlaHisMetLeuAsnTrpLeuArgPhePh 191
 301 CTCAACCGCAAGCTCTATCACTGTGTATGTATGTATGTATGTATGTATGT 350
 191 eProLeuGlyHisIleHisIleValAspGlyAspArgLeuValArgAsp 208
 351 CCGGTGGCCACATCCACATTTGGAGCGGACCGCTCATAGGAGCC 400
 208 rPheProGluIleGlnLysValGluArgPheLeuLysLeuSerProGln 224
 401 CCTTCCTCGATCCAAAGGTCAGAGGTTCCCTAAAGCTGTCGGCCAC 450
 225 IleAsnAlaSerAsnPheTyrPheAsnLysThrLysGlyPheTyrCysLe 241
 451 ATCAATGCTCGAATCTCTATCTTAAACAACAACCAAGGCTTTTACGGCT 500
 241 uArgAspSerGlyLysAspArgCysLeuHisGluSerLysGlyArgAlaH 258
 501 GCGGACAGCGCGCGGAGCGTGTGTACATAGTCCAAAAGAGCGTGGCC 550
 258 isProGluValAspProLysLeuAspLysLeuHisGluTyrPheHis 274
 551 ACCCCCAAGTCATCCCAAACTACICATAAATACGACGAATAGTTTCAT 600
 275 GluProAsnLysLysPhePheLysLeuValGlyArgThr 287
 601 GAGCCATAAAGATGTTCTCGAGCTTGTCGAGCAACA 639

seq_name: gb_est1:A1870763

seq_documentation_block:

LOCUS A1870763 609 bp mRNA linear EST 07-MAR-2000
 DEFINITION w176101.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2430841 3' similar to TP-014792 G14792 HEPARAN SULFATE 3-O SULFOTRANSFERASE-1 PRECURSOR ;, mRNA sequence.
 ACCESSION A1870763

```

VERSION      A1870763.1  GI:5544731
KEYWORDS     EST.
SOURCE       Homo sapiens
ORGANISM     Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
              Mammalia; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE    1 (bases 1 to 609)
AUTHORS     NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap/
TITLE       National Cancer Institute / National Institute of Neurological
              Disorders and Stroke, Brain Tumor Genome Anatomy Project
              (CGAP/BTCAP), Tumor Gene Index
JOURNAL      Unpublished (1998)
COMMENT      Contact: Robert Strausberg, Ph D
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfield M.D.,
              Ph.D.
              CDNA Library Preparation: M. Bento Soares, Ph D, M Fatima
              Bonaldo, Ph.D.
              CDNA Library Arrayed by: Greg Isomura, Ph.D.
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I M A G E Consortium/LLNL at:
              www.bio.llnl.gov/bbrp/image/image.html
              Insert length: 1024  Std Error: 0 00
              Seq primer: -400p from Gibco
              High quality sequence stop: 451.
FEATURES     Location/Qualifiers
              1..609
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone_lib="IMAGE:2430841"
              /clone_lib="NCI_CGAP_Brn25"
              /tissue_type="anaplastic oligodendrolioma"
              /lab_host="pH10B"
              /note="Organ: brain; Vector: pTT73D-Pac (Pharmacia) with a
              modified polylinker, Site_1, Not I, Site_2, Eco RI; 1st
              strand cDNA was primed with a Not I + oligo(dT) primer 15'
              TGTACCAATCGAATGCGAGCGCGCCATAGGTTTTTTTTTTTTTTTTTTT
              T 3'; double-stranded cDNA was ligated to Eco RI
              adaptors (Pharmacia), digested with Not I and cloned into
              the Not I and Eco RI sites of the modified pTT73 vector
              library is normalized, and was constructed by Bento
              Soares and M.Fatima Bonaldo."
BASE COUNT   144 a 201 c 148 g 115 t 1 others
ORIGIN
alignment_scores
  Quality: 1024.00      length: 202
  Ratio: 5.172         Gaps: 0
  Percent Similarity: 98.020  Percent Identity: 90.594
alignment_block
  US-09-557-262_2_COPY_21_307 x A1870763
  Align seq 1/1 to: A1870763 from: 1 to: 609
75  AspTrpGluGluHisTyrSerGlnGlyLeuGlyTTPtyrLeuTyrGlnMe 91
1  CACTCGGAGGAGCATACACCGCAGCGCTTGGCTTGGTACCTCAGCCAGAT 50
91  tProPheSerSerProHisGlnLeuThrValGluTysThrProAlaTyrP 108
51  GCGCTTCTCTAGGCAAAAGAGGATATATGTAAGAGAGAGAGAGAGATT 100
108  heThrSerProTyrValProGluArgGluHisSerMetAsnProThrIle 124
101  TCACCTGCGCCCAAGAGCGCGACAGGAGATACAGCAIAGAACCGGTCATC 150
125  ArqLeuLeuLeuLeuLeuArqAspProSerGluArqValLeuSerAspTy 141
151  CGGCTGCTGCTTCATCCCTGGCAGACCGCGCGAGCGCGTGGTATCGACTA 200

```

```

141  rThiGluValLeuTyrAsnHisLeuGluGlyHisTyrGlyTyrProPro 158
201  CACTCAATGCTGTATACAAATACATATAGAAATATAGATCTATGCGTCA 250
158  LeGluSerLeuLeuMetAlaAspGlyArgLeuAsnLeuAspTyrTyrAla 174
251  GGTATATATATATGCTGTCACGCGAGGCTCAATGTCGATACAAAGGCG 300
175  LeuAsnArqSerLeuTyrHisAlaHisMetLeuAsnTrpLeuArqPhePh 191
401  CTCAATGAGAGATCTATATATGTAATATGTAATATGTCGTCGCTCTT 450
191  eProLeuGlyHisHisHisValAspGlyAspArgLeuLeuArqAspP 208
351  CCGCTGAGGATATATATATATGTAATATGTAATATGTCGTCGCTCTT 400
208  PheProGluGluGlnValGluValGluArqPheLeuLeuSerProGln 224
401  CTTTCTCTGACATCTAAAGAGCTGACAGGCTTCTTAAAGCTGTCGCGCAG 450
225  IleArqAlaSerAspPheTyrPheAsnLysThrLysGlyPheTyrCysLe 241
451  ATCAATCTTCTCAATCTTCTTAAATTAATTAATTAATTAATTAATTAAT 500
241  GATGASpSerClyLysAspArgCysLeuHisGlySerLysClyArgAlaH 258
501  GCGGATATATGTAATATGTAATATGTAATATGTAATATGTAATATGTA 550
258  IsProGlnValAspProLysLeuLeuAspLysLeuHisGluTyrPheHis 274
551  ACCCGCAAGCTGACATCCCAAACTACTAATAAATGCGACCAATATTTTCA 600
275  GluPro 276
601  GAGCCA 606
seq_name: gb_est2:BC820479
seq_documentation_block
  LOCUS      BC820479
  DEFINITION  502782294F1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:4933676
  ACCESSION  BC820479
  VERSION    BC820479.1  GI:14168066
  KEYWORDS   EST.
  SOURCE     human.
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
              Mammalia; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE   1 (bases 1 to 788)
AUTHORS     NIH-MGC http://mgc.ncbi.nlm.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph D.
              Email: cgapbs-remail.nih.gov
              Tissue Procurement: David N. Louis, M.D.
              CDNA Library Preparation: Life Technologies, Inc.
              CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM10861 row: q column: 21
              High quality sequence stop: 777.
              Location/Qualifiers
              1..788
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone_lib="IMAGE:4933676"
              /clone_lib="NCI_CGAP_Brn25"
              /tissue_type="anaplastic oligodendrolioma with 1p/19q
              loss"
              /lab_host="DH10B (11 phage-resistant)"

```


/note="Organ: brain; Vector: pCMV-Sport6; Site: 1 Not L
 Site: 2; SalI, cloned unidirectionally. Primer oligo: dl
 Average insert size 2.3 kb constructed by Life
 Technologies. Note: this is a NCI-CGAP library"

BASE COUNT 209 a 224 c 174 g 181 t

ORIGIN

alignment_scores:

Quality: 1019.00 Length: 202
 Ratio: 5.145 Gaps: 0
 Percent Similarity: 98.020 Percent Identity: 91.089

alignment_block:

US-09-557-262-2_COPY_21_307 x BG820479 ..

Align seq 1/1 to: BG820479 from: 1 to: 788

```

86 TrpTyrLeuThrGlnMetProPheSerProHisGlnLeuThrValGI 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3 TGGTACCTCAGCCAGATGCTCTCTCTGGGACACACAGTACACAGTGA 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
102 uLysThrProAlaTyrPheThrSerProLysValProGluAroIleHis 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
53 GAAGACCCCGCGTATTTCACGTCGCCCAAGAGCCCHCAGCGAGCTACA 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
119 erMetAsnProThrIleArgLeuLeuLeuLeuLeuArgAspProSerGlu 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
103 GCATCAACCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
136 ArgValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLysH 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 CGGTGCTATGTAATGACTAAGCCAAATGTTTACAAACACATGAGAAAT 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
152 slsProTyrProProIleGluAspLeuLeuMetArgAspGlyArgLeuA 169
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
203 CAAGCCGACCCGCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
169 snLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisAlaHisMetLeu 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
253 ATGTGGATACAAAGCCCTCAACGCGAGCCTCTACACGCTGACATGAG 302
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
186 AsnTyrLeuArgPhePheProLeuGlyHisIleHisIleValAspGly 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
303 ACTGGCTGCGCTTTTTCGCGCTGCGCTGCGCTGCGCTGCGCTGCG 352
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
202 pArgLeuLeuArgAspProPheProGluIleGlnLysValIleuArgPhe 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353 CCGCCTCATCAGGACCCCTTCCCTGAGATCCAAAGGTCGAGAGGTTC 402
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
219 euLysLeuSerProGlnIleAsnAlaSerAsnPheTyrPheAsnLysThr 235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
403 TAAAGTTPGCGAGAGAGATATATGTAATGTAATGTAATGTAATGTA 452
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
236 LysGlyPheTyrCysLeuArgAspSerClyLysAspArgCysLeuHisGI 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
453 AAGGCTTTTATATATATATATATATATATATATATATATATATAT 502
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
252 uSerLysGlyArgAlaHisProGlnValAspProLysLeuLeuAspLysL 269
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
503 GTCCAAAGCCCGCGCGCCACCCCAAGTCGATCCCAAACTACTCAATAA 552
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
269 euHisclutyrPheHisGluProAsnLysLysPhePheLysLeuValGly 285
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
553 TGAATGATATTTTCATAGGCTTAATAAGAAAGTTCTTCGAGCTTGTC 602
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
286 ArgThr 287
|||||
603 AGACA 608

```

seq_name: qb_est2.BE676274

seq_documentation_block:

HE676274 1 621 bp mRNA (cloned) EST 08-SEP 2000
 7724322 x1 NCI-CGAP-CL1 Homo sapiens cDNA clone IMAGE:3295779 37
 similar to U014792.014792 HEPAPAN SULFATE 3 c SWIFTTPANSEPERASE-1
 PRECURSOR. 17 mRNA sequence.
 BE676274 1 621 bp mRNA sequence.
 HE676274 1 621 bp mRNA sequence.
 EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 621)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
 JOURNAL Tumor Gene Index
 COMMENT Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 EMAIL: cgaps@mail.nih.gov
 TISSUE Procurement: Ash Alizadeh, John Hyrd, M.D., Mike Grever,
 M.D., Louis M. Staudt, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/URL, send email to:
 infoimage.lnl.gov
 Seq primer: -400P from Gibco
 High quality sequence stop: 383.
 FEATURES
 Location/Qualifiers
 1..621
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3295779"
 /clone_lib="NCI-CGAP-CL1"
 /tissue_type="B-cell, chronic lymphocytic leukemia"
 /lab_host="H108"
 /note="Vector: pT73D Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5'
 TGTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTT
 T 3']; double-stranded cDNA was ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of the modified pT73 vector.
 Library is normalized, and was constructed by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 141 a 204 c 150 g 125 t
 GPGIN

alignment_scores:

Quality: 1007.00 Length: 207
 Ratio: 5.060 Gaps: 0
 Percent Similarity: 96.145 Percent Identity: 86.957

alignment_block:

US-09-557-262-2_COPY_21_307 x BE676274 ..

Align seq 1/1 to: BE676274 from: 1 to: 621

```

71 ValHisPhePheAspTyrGluGluHisTyrSerGlnGlyLeuGlyTyrI 87
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 GTCCACTTCTTCGACTGGGAGGACCATACAGCCAGCTTGGCGTGTGA 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 GluThrGlnMetProPheSerSerProHisGlnLeuThrValHisLysT 104
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 CTCACGCCAGATGGCGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
104 hrProAlaTyrPheThrSerProLysValProGluArgIleHisSerMet 120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
101 CCGCGGCTATTTTCATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 150
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 AsnProThrIleArgLeuLeuLeuIleLeuArgAspProGluArgVa 137
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```


info@image.llnl.gov

Seq primer: -400p from Gibco

High quality sequence stop: 474

Location/Qualifiers

FEATURES

source

1..612

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3648661"

/clone_lib="NCL_CGAP_Pr28"

/sex="male"

/dev_stage="adult"

/lab_host="DH10B"

/note="Organ. prostate. Vector. p1713b-Pac (Pharmacia)

with a modified polylinker. Plasmid DNA from the normalized library NCL_CGAP_Pr22 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 985608-986759, 1101192-1101959, and 1217928-1220615). Subtraction by Benito Soares and M. Fatima Bernaldo.

BASE COUNT 138 a 202 c 150 g 121 t 1 others
ORIGIN

alignment_scores:

Quality: 981.50 Length: 214
Ratio: 5.033 Gaps: 1
Percent Similarity: 91.121 Percent Identity: 84.112

alignment_block:

US-09-557-262-2_copy_21_307 x BF447174

Align seq 1/1 to: BF447174 from: 1 to: 612

71 ValHisPhePheAspTrpGluGluHisLysSerGlyLeuGlyLeuGlyTrp 87
|||||
1 GGCACCTTCCTCCAGCTGGAGAGCATACAGCCACGGCTTGGCTGGCA 50
87 rleuThrGlnMetProPheSerProHisGlnLeuThrValGluLys 104
|||||
51 CCTCAGCCAGATGCTTCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 100
104 hrProAlaTyrPheThrSerProLysValProGluArgIleHisSerMet 120
|||||
101 GCGGGGCTATTCAGTCGAGCAAAAGTGGCTGAGCGAGTCTACAGCATG 150
121 AsnProThrIleArgLeuLeuLeuLeuArgAspProSerGluArgVa 137
|||||
151 AACCCGTGCAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 200
137 leuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLysHisLysP 154
|||||
201 GCTATCTGACTACACCAAGTGTCTTACAACTCAATGCAAGCAACCAAG 250
154 rotyProProIleGluAspLeuLeuMetArgAspGlyArgLeuAsnLeu 170
|||||
251 GCTAGGCTTATGAGAACTTCTGATGAGAGAGAGAGAGAGAGAGAGAG 400
171 AspTyrLysAlaLeuAsnArgSerLeuTyrHisAlaHisMetLeuAsnTr 187
|||||
301 CACTACAGGCGCTCAACCGCAGCTTACACAGCTGCTGCTGCTGCTGCTG 350
187 pLeuArgPhePheProLeuTyrHisIleHisTleValAspLysAspArg 404
|||||
351 GCTGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 400
204 cgluThrArgAspThrPheProGlnIleGlnLysValIleuArgPheLeuLys 220
|||||
401 TCATAGGAG 450
221 leuSerProGlnIleAsnAlaSerAspPheTyrPheAsnLysThrLysG 237
|||||

451 CTGTGCGCGCAGCATCAATGCT..... 471
237 yPheTyrCysLeuArqAspSerGlyLysAspArqCysLeuHisGluSerL 254
|||||
472 TATGCTGTGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 517
254 YSGLYARGAlaHisProGlnValAspProLysLeuLeuAspLysLeuHis 270
|||||
518 AANAGATGAG 567
271 GluTyrPheHisGluProAsnLysLysPhePheLysLeuVal 284
|||||
568 CAATATGTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 609

seq_name: qb_est2:BG746348

seq_documentation_block:

LOCUS BG746348 783 bp mRNA linear EST 15-MAY 2001
DEFINITION 602703607F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4857025 5',
mRNA sequence.

ACCESSION BG746348 1 GI-14057001

VERSION BG746348

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 783)

AUTHORS NIH-WHS' http://imgc.ncbi.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs.rem@nih.gov

Tissue Procurement: ATCC

CDNA library Preparation: Ling Hong/Kubin Laboratory

CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: NIH Intramural Sequencing Center

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM1708 row: n column: 02

High quality sequence stop: 720.

FEATURES

source

1..783

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4857025"

/clone_lib="NIH_MGC_15"

/tissue_type="adenocarcinoma cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ. colon; Vector: pUTB7; Site_1: XhoI; Site_2:

EcoRI; cDNA made by oligo-dt priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGACGAG(C) Size-selected, 500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT 181 a 255 c 176 g 171 t

ORIGIN

alignment_scores:

Quality: 944.50 Length: 222
Ratio: 4.563 Gaps: 7
Percent Similarity: 93.243 Percent Identity: 83.784

alignment_block:

US-09-557-262-2_copy_21_407 x BG746348

Align seq 1/1 to: BG746348 from: 1 to: 783

73 PhePheAspTrpGluGluHisTyrSerGlnGlyLeuGlyTrpLeuLeu 89

```

|||||
3  TTCTTGAGTGGAGGAGGATTACAGCAGCGCTTGAGCTGGTACCTCAG 52
89  rClnMetProPheSerProHisGlnLeuThrValGlnLysThrProA 106
53  CAGAGAGGCTTCTGTTGAGTAAAGTAAAGTAAAGTAAAGTAAAG 102
106  laTyPheThrSerProLysValProGluArgIleHisSerMetAspPro 122
103  CGTATTTCAGCTGCGCCAAAGTGGCTGAGCGAGTCTACAGCATGAACCG 152
123  ThrIleArgLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 139
153  TCATTCGGGCTGCTGTTATGTTGAGAGGAGGAGGAGGAGGAGGAG 202
139  rAspTyThrGlnValLeuTyAsnHisIleGlnLysHisLysProTyP 156
203  TGACTACACCCGAGTGTCTACACCCGATGCAGACAGCAGCCCTACC 252
156  roProIleGluAspLeuLeuMetArgAspGlyArgLeuAsnLeuAspTy 172
253  CCTCATTCAGCAGGATTCCTGCTGCGGATGCCAGGCTCAATGTGGACT 302
173  LysAlaLeuAsnArgSerLeuTyHisAlaHisMetLeuAsnTrpLeuAr 189
303  AAGGCCCTCAACCGAGGCTCTACACGCTGACATGAGAGAGTGGCTGG 352
189  gPhePheProLeuGlyHisIleHisIleValAspGlyAspArgLeuIle 205
353  CTTTTTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 402
206  ArgAspProPheProGluIleGlnLysValGlnArgPheLeuLysLeu 221
403  AGGGAGCCCTTCCTCCGAGATCAAAATGCTGAGAGGTTCTAAATGCTG 452
222  SerProGlnIleAsnAlaSerAspPheTyPheAsnLysThrLysGlyP 238
453  TCGCCGAGATTAATGCTGAGATTTCTTCTTAAACAAACCAATGAGCT 502
238  heTyrcysLeuArgAspSerGlyLysAspArgCysLeuHisGluSerLy 254
503  TTACTGCTGGGAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 552
254  s.GlyArgAlaHisProGlnValAspProLysLeuLeuAspLysLeuHis 270
553  CAGGCGGGGCGGACCGCCGAGGATGCTGCTGCTGCTGCTGCTGCTG 602
271  GluTyPheHisGluPro...AsnLysLysPhePheLysLeuValGlyAr 286
603  GAATATTCATGAGCCCATATAGAGAGTTTTCGACAGCTTGTTGAGAG 652
286  gthr 287
653  AAC 656

```

seq_name: qb_est1.A1431882

seq_documentation_block:

LOCUS A1431882 570 bp mRNA linear EST 30-MAR-1999
 DEFINITION U26C09 x1 NCI-CGAP_K1311 Homo sapiens cDNA clone IMAGE:2131600 47
 similar to TR-014792 014792 HEPARAN SULFATE 4-O-SULFOTRANSFERASE-1
 PRECIPITOR .. mRNA sequence

ACCESSION A1431882

VERSION A1431882.1 GI:4306292

KEYWORDS EST.

SOURCE human

ORGANISM Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 570)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

JOURNAL COMMENT

Unpublished (1997)

Contact: Robert Strausberg, Ph.D

Email: cgaps@mail.nih.gov

Tissue Procurement: Christopher Miskaluk, M.D., Ph.D., Michael K.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/Li.NI.ac:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length 621 Std Error 0.00

Seq primer: -40mp from 3' end

High quality sequence stop: 457.

FEATURES

SOURCE

1..570
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NCI-CGAP_K1311"
 /lab_host="DH10B"
 /note="Organ: kidney; Vector: pUT30-Pac (Pharmacia) with
 a modified polylinker; Site: i. Not i. Site: 2. Eco RI;
 Plasmid DNA from the normalized library NCI-CGAP-K131 was
 prepared, and ss circles were made in vitro. Following HAP
 purification, this DNA was used as a tracer in a subtractive
 hybridization reaction. The driver was PCR amplified cDNAs
 from a pool of 5,000 clones made from the same library
 (clone IDs 1322376-1323911, 1456007-1456775, and
 1460552-1501855). Subtraction by Ron's Soares and M.
 Fatima Bonaldo."

BASE COUNT 124 a 130 ~ 145 g 110 t 1 others
 ORIGIN

alignment_scores:

Quality: 934.00 Length: 190

Ratio: 5.104 Gaps: 0

Percent Similarity: 96.316 Percent Identity: 88.947

alignment_block:

US-09-557-262-2_COPY_21_307 x A1431882

Align seg 1/1 to: A1431882 from: 1 to: 570

75 AspTrpGluGluHisTySerGlnGlyLeuGlyTrpTyLeuThrGlnMe 91
 1 GATTTGGAGAGATTAATAGATAGAGTTGGATGTTATATAGCCAGAT 50
 91 tProPheSerProHisGlnLeuThrValGlnLysThrProAlaTyP 108
 51 GCGGCTGCGGAGACAGACAGACAGACAGACAGACAGACAGACAG 100
 108 heThrSerProLysValProGluArgIleHisSerMetAspProThrIle 124
 101 TATATGAGGAAAATGATGATGATGATGATGATGATGATGATG 150
 125 ArgLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 141
 151 CGGTGATGTTATGTTGATGTTGATGTTGATGTTGATGTTGATG 200
 141 rThrGlnValLeuTyAsnHisIleGlnLysHisLysProTyPProI 158
 201 CACCCCAAGTGTCTACACCCATGACAGACAGACAGACAGACAG 250
 158 leGluAspLeuLeuMetArgAspGlyArgLeuAsnLeuAspTyLysAla 174
 251 TGGAGAGATTTGATGAGGAGATGAGGAGATGAGGAGATGAGG 300
 175 LeuAsnArgSerLeuTyHisAlaHisMetLeuAsnTrpLeuArgPhe 191
 301 CTAAACGGTAGGTTTACAGAGTGCATAGAGAGAGAGAGAGAGAG 350


```
alignment_scores
      Quality: 1544.00      Length: 287
      PctIdent: 5.580      Gaps: 0
      Percent Similarity: 100.000
```

Align seq 1/1 to: AAX37245 from: 1 to: 1685

US-09-557-262-2 COPY 21 397 X AAX37246

Align seq 1/1 to: AAX37246 from: 1 to: 1305

```

2  ProAlaAlaProGlyProGlyLeuLeuGlnGlnLeuLeuArgLysVa 18
   |||||
173 CTTCTCCCGCCCGCCGAG CTAGCCACAGCAGAGCTTTCTGCGGAAAGC 219
   |||||
18 IfleleleuProGluAspThrGlyGlyGlyThrAlaSerAsnGlySert 35
   |||||
220 GGGGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 264
   |||||
35 hrGlnGlnLeuProGlnThrIleIleIleGlyValArgGlyGlyThr 51
   |||||
270 CCGACAGTTCGCCGACACCATCATCATCGCCCTGCCCAAGCGCGCAGC 319
   |||||
52 ArgAlaLeuLeuGlnMetLeuSerLeuHisProAspValAlaAlaGln 68
   |||||
320 CCGCCACTTCGCGACATGCTCAGCCTGCACCCGACCTGCCGCGCCGCA 369
   |||||
68 uAsnGluValHisPheAspTrpGluGluHisTyrSerGlnGlyLeuG 85
   |||||
370 GAAAGAGTTCATTTCTGAGTGGAGAGGAGATTAAGGACAGGCTTG 419
   |||||
85 IyTrpTyrLeuThrGlnMetProPheSerSerProHisGlnLeuThrVal 101
   |||||
420 GCTGTACTCTAGCCAGATGCCCTTCCTCGGCCACACACAGCTCACAGT 469
   |||||
102 GluLysThrProAlaIyPheThrSerProLysValProGluArgIleHis 118
   |||||
470 GAGAGACCCCGCGGTATTTCCAGCTGCCCAAGTGCCTGAGCGACICTA 519
   |||||
118 sSerMetAsnProThrIleArgLeuLeuLeuLeuArgAspProSerG 135
   |||||
520 CAGCATGAAGAGTGCATCGAGCTGATCTCTATCTGCGAGAGCGGTCG 569
   |||||
135 IuArgValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLys 151
   |||||
570 ACCGCGCTGCTATCTGACTACACCAAGTGTCTACACACCATGCCAAG 619
   |||||
152 HisLysProTyrProProIleGluAspLeuLeuMetArgAspGlyArgLe 168
   |||||
620 CACAAGCCCTACCGCTCCATCGAGGACATCTCTGCTGCGCGCATGCGC 664
   |||||
168 uAsnLeuAspTyrLysAlaLeuAsArgSerLeuTyrHisAlaHisMet 185
   |||||
670 CAATGTGATACAAAGGAGGATTAAGAGAGGCTTATTAAGGTGACATGC 719
   |||||
185 euAsnTrpLeuArgPhePheProLeuGlyHisIleHisIleValAspGly 201
   |||||
720 ACAACTCGCTCGCTTTTCCCGCTCCCGCCACATCCACATCTCGACCGC 769
   |||||
202 AspArgLeuIleArgAsnProPheProGluIleGlnLysValGluArgPhe 218
   |||||
770 GACCGCTCATCAGCGACCGCTTCCCTGACATCCAAAACGCTCCGAGCTT 819
   |||||
218 eleuLysLeuSerProGlnIleAsnAlaSerAsnPheTyrPheAsnLysT 235
   |||||
820 CTTAAGTGTGAGGAGATTAAGATTTTAAATTTTAAATTTTAAATTTTAA 864
   |||||
235 hrLysGlyPheTyrCysLeuArgAspSerGlyLysAspArgCysLeuHis 251
   |||||
870 CCAAGGCTTTTACTGCTGCGGACACAGCGCGCGGACCGCTGCTTACAI 919
   |||||
252 GluSerLysGlyArgAlaLysProGlnValAspProLysLeuLeuAspLys 268
   |||||
920 GACTCCAAAGCGCGCGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCG 969
   |||||
268 sLeuHisGluTyrPheHisGluProAsnLysLysPhePheLysLeuValG 285
   |||||
970 ACTGACGAATATTTTATGAGATTAATAGAGATTTTGAAGTTCGTTT 1019
   |||||
285 IyArgThr 287
   |||||

```

1020 GCAGAAC 1027

seq_name: /STD5/qcdata/geneseq/geneseq-emb1/NA2000.DAT-AAZ36687

seq_documentation_block:

ID AAZ36687 standard; DNA; 1305 BP.

XX AAZ36687;

XX 14 MAP-2000 (first entry)

XX Nucleotide sequence for GenBank accession number AF019386.

XX Stimulus-regulated nucleic acid; sequence profile; nucleic acid level;
 XX differentially expressed nucleic acid; disease state; cancer;
 KW autoimmune disease; infectious disease; aging; developmental disorder;
 KW proliferative disorder; neurological disorder; toxicity;
 KW treatment resistance; differential expression; drug discovery;
 KW growth factor; epidermal growth factor; radiation; stress; pathogen; ss.

OS Homo sapiens.

XX W0955913-AZ.

XX 04-NOV-1999.

XX 27-APR-1999; 99W0-US09119.

XX 27-APR-1999; 98US-0084331.

XX 27-AUG-1998; 98US-0098070.

XX 04-FEB-1999; 99US-0118624.

XX (KIMM-) KIMMEL CANCER CENT SIDNEY.

XX McClelland M, Welsh J, Trenkle T;

XX WPI; 2000-086388/07.

XX Measuring expression of low abundance reduced complexity target nucleic acid molecules -

XX Disclosure; Fig 14; 137pp; English.

XX AAZ36687-236725 represent nucleotide sequences from stimulus regulated nucleic acid molecules. The sequences represent a profile of sequences which can function as targets in the method of the invention. The specification describes a method for measuring the level of two or more nucleic acid molecules in a target. The method comprises contacting a probe with an arbitrarily or statistically sampled target and detecting the amount of specific binding of the target to the probe. The methods can be used to identify differentially expressed nucleic acid molecules associated with disease states, such as cancer, autoimmune disease, infectious disease, aging, developmental disorder, proliferative disorder or neurological disorder. Alternatively the methods can be used to assess the efficacy of toxicity of or a resistance to a treatment. Also the methods can be used to determine differential expression of nucleic acid molecules in response to a stimulus, e.g. a chemical, drug or growth factor (especially epidermal growth factor), radiation, stress or a pathogen. The methods can also be used to determine over-regulated genes that can be potential targets for drug discovery.

XX Sequence 1305 BP; 304 A; 380 C; 329 G; 292 T; 0 other;

alignment_scores:

Quality: 1345.50 Length: 286

Ratio: 4.928

Percent Similarity: 95.455 Percent Identity: 87.413

alignment_block:

US-09-557 262 2_COPY_21_307 x AAZ36687


```

754 CCCACACTGCCCCACACATATATATATCGGCGGCGGAAAGCGCGGACG 705
52 ArqAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaAlaG1 68
704 CCGCCACTGCTGTATATATATATATATATATATATATATATATAT 655
68 uAsnGluValHisPhePheAspThrGluGluHisTyrSerGlnCylLeuG 85
654 GAAACAGATTAATTTTAAATTAAGAGAGAGATTAATTAAGAGATTA 605
85 TyrPtyrLeuThrGluMetProPheSerSerProHisGlnLeuThrVal 101
604 CCGTGTATATATATATATATATATATATATATATATATATATATAT 555
102 GlnLysThrProAlaTyrPheThrSerProLysValProGlnArgGln 118
554 CAGACAGCCCGGCTATTTATATATATATATATATATATATATATAT 505
118 sSerMetAsnProThrIleArgGlnLeuLeuIleLeuArgAspProSerG 135
504 CAGCATAAATGAGATATATATATATATATATATATATATATATATAT 455
135 IuArgValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLys 151
454 AGCGGCTGTATATATATATATATATATATATATATATATATATATAT 405
152 HisLysProTyrProProIleGlnAspLeuLeuMetArgAspGlnArg 168
404 CACAGGCTTACCGTATATATATATATATATATATATATATATATATAT 355
168 uAsnLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisAlaHisMet 185
354 CAACTGTGACTACAAAGGCTTAAATATATATATATATATATATATAT 305
185 euAsnThrLeuArgPhePheProLeuGlnCysHisIleHisIleValAspCly 201
304 ACAACTGCTGCGGCTTTTCCCGCTCGCCACATCCACATTCGACGGC 255
202 AspArgLeuIleArgAspProPheProGluIleGlnLysValGlnArgPh 218
254 CACCCCTTCATCAGCGACCCCTTCCCTGCATCCCAAAAGTCGACAGCT 205
218 GlnLysLeuSerPtyrGlnIleAspAlaSerAsnProPheTyrPheAsnLys 235
204 CATTAAATATATATATATATATATATATATATATATATATATATAT 155
235 hriLysGlyPheTyrCysLeuArgAspSerGlyLysAspArgCysLeuHis 251
154 CCAAGCGTATATATATATATATATATATATATATATATATATATAT 105
252 GluSerLysGlyArgAlaHisProGlnValAspProLysLeuLeuAsp 268
104 CAGTCCAAAGCGCGCGGACCCCAAGTCCATCCCAAACTACTCAATAA 55
268 sLeuHisClnLyrPheHisClnProAsnLysLysPhePheLysLeuValC 285
54 ACTGATATATATATATATATATATATATATATATATATATATATAT 5
285 Ty 285
4 GC 3

```

seq_name: /SDS5/seqdata/geneseq/geneseq-emb1/NA2001A.DAT:ABA35017

seq_documentation_block:

ID ABA35017 standard: IMA: 924 bp

XX

AC

XX

XX

DT

XX

DE

21-JAN-2002 (first entry)

Probe #13483 for gene expression analysis in human heart cell sample

XX Human; gene expression; heart; microarray; vascular system; probe;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease; ss.
 XX Homo sapiens.

XX WO200157274-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000666.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 09-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0643666.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0924263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human heart

XX Claim 4; SEQ ID NO 13483; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart. The present sequence is one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.

XX Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from Wipo at http://wipo.int/pub/published_pat_sequences.

XX SQ Sequence 924 BP, 169 A, 252 C, 310 G, 193 T; 0 other;

alignment_scores:

Quality: 1435.50 Length: 284

Ratio: 4.928 Gaps: 1

Percent Similarity: 95.423 Percent Identity: 87.324

alignment_block:

US-09-557-262-2_COPY_21_307 x ABA35017/rcv

Align seq 1/1 to reverse of: ABA35017 from: 1 to: 924

2 ProAlaAlaProGlyProGlyLeuLysGlnGlnGlnLeuLeuArgLysVa 18

851 CCGTCCCGCGCGCGCGGAC CTAGCCACGACGACGCTTCGCGCAAGC 805

18 LilelleLeuProGluAspThrGlyGluThrAlaSerAsnGlySerT 45

804 GGGAAATTAACAGATGATGATGATGATGATGATGATGATGATGATG 755

35 hGlnGlnLeuProGlnThrIleIleIleIleIleIleIleIleIleI 51

754 GCGAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 705

52 ArgAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaAlaG1 68

704 CCGCCACTGCTGTATATATATATATATATATATATATATATATAT 655

```

68  uAsnGlnValHisPhePheAspTrpGlnGlnHisTyrSerGlnIndLeuArg 85
654 GAACGAGGTCACCTCTCTCCACCTGGGAGGAGCATACAGCCACGCGTCG 605
85  LyrTrpYrLeuThrGlnMetProPheSerSerProHisGlnLeuThrVal 101
604 GCTGGTAACTTCAAGAAAGATGGTCTCTCTGGAGAAAGAAATATAT 555
102 GlnLysThrProAlaLysThrThrSerProLysValTrpGluArgLeu 118
554 GAAAGAAAGAGAGGATATTTATATGATGAGGAAAGAGATGATATATA 505
118 sSerMetAsnProThrIleArgLeuLeuLeuIleLeuArgAspProSer 135
504 CACCATGCAACTGGTTCATCCGCGCTGCTGTCATGCTGGTAGACCTCG 455
135 LuArgValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLys 151
454 AGAGAGATGTATTTATTTATATATCAAGATGTTTATATATATATGAG 405
152 HisLysProTrpProThrIleGlnAspLeuLeuMetArgAspGlnArg 169
404 CATAAGGCTATACCGCTCATCGAAGAGATGTTCTGGTTCGATGAGAG 355
168 uAsnLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisAlaHisMet 185
354 CAATGTCGACTACAGCCCTCAACCCGACGCTCATCCACGTCACATCC 305
185 euAsnTrpLeuArgPhePheProLeuGlyHisIleHisIleValAspGly 202
304 AGAACTGCTGGGCTTTTTCGGCTGGAGTAATATATATTTTACACGAG 255
202 AspArgLeuIleArgAspProPheProGluIleGlnLysValGluArgP 218
254 GACCGCTATATAGGACACCGCTTCCTGGATCCCAAAAGTCGAGAGGT 205
218 eLeuLysLeuSerProGlnIleAsrAlaSerAsnPheTyrPheAsnLys 235
204 CCTTAAACGTGTCGCCGACATCAATGCTTCGAACTTCCTACTTTTACAAA 155
235 hrLysGlyPheTyrCysLeuArgAspSerGlyLysAspArgCysLeuHis 251
154 CCAAGGCTTTTACTGCTGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAG 105
252 CusSerLysGlyArgAlaHisTrpGlnValAspProLysLeuLeuAsp 268
104 GATCTCAAAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 55
268 sLeuHisGluTyrPheHisGluProAsnLysLysPhePheLysLeuVal 285
54 ACTGCAGCAATATTTTCATGAGCTCAATTAAGAAGATGCTTCGACAT 5
285 1y 285
4 GC 3

```

seq_name: /SIDS5/4c4data/4c4neseq/4c4neseqnemb1/NA2007A.DA...AAK16371

seq_documentation_block:

ID AAK16371 standard; DNA; 924 bp.

XX
XX

AC AAK16371;

XX
XX

DT	05-NOV-2001 (first entry)
XX	
DE	Human brain expressed single exon probe SEQ ID NO: 16362.
XX	
KW	Human; brain expressed exon, gene expression analysis; probe;
KW	microarray; Alzheimer's disease, multiple sclerosis, schizophrenia;
KW	epilepsy; cancer; ss.
XX	
OS	Homo sapiens.


```

168 uAsnLeuAspTyrHisAlaLeuAsnArgSerLeuTyrHisAlaHisMetL 185
|||||
354 CAATATGCAATACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 405
|||||
185 euAsnTrpLeuArgPhePheLeuGlyHisLeHisLeValAspGly 201
|||||
304 AGAACTGGCTGGCTTTTTCGGCTGGCTGGCTGGCTGGCTGGCTGGCT 355
|||||
202 AspAlaLeuLeuArgPheProGlnLeuGlnLysValGluArgPhe 218
|||||
254 GACGCGCTGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 305
|||||
218 cLeuLysLeuSerProGlnLeuAsnAlaSerAsnPheTyrHisAsnLys 235
|||||
204 CCAAGAGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 255
|||||
235 hLysGlyPheTyrCysLeuArgAspSerGlyLysAspArgCysLeuHis 251
|||||
154 CCAAGAGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 205
|||||
252 GluSerLysGlyArgAlaHisProGlnValAspPheCysLeuAspGly 268
|||||
104 GAGTCCAAAGGCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 155
|||||
268 sLeuHisGluTyrPheHisGluProAsnLysLysPhePheLysLeuValG 285
|||||
54 ACTGCACGAATATTTTCATGAGGCAATATGAGGCAATATGAGGCTG 5
|||||
285 ly 285
||
4 GC 3

```

seq_name: /SID55/grqdata/geneseq/geneseq-emb1/NA2001A.DAT:AA148186

seq_documentation_block:

ID AA148186 standard; DNA; 924 bp.

XX AC AA148186;

XX DT 17-OCT-2001 (first entry)

XX DE Probe #16872 used to measure gene expression in human placenta sample.

XX KW Probe; microarray; human; placenta; antenatal diagnosis;

XX KW genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200157272-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00663.

XX PR 04-FEB-2000; 2000US-0180412.

XX PR 26-MAY-2000; 0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GH-0024253.

XX PA (MOLE*) MOLECULAR DYNAMICS INC.

XX XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX XX WPI; 2001-488897/53.

XX XX Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human placenta -

XX XX Claim 25; SEQ ID NO 16872; 654bp; English.

XX

CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.

XX

50 Sequence 924 BP; 169 A; 252 C; 310 G; 193 T; 0 Other.

alignment_scores:

Quantity: 1335.50 Length: 284

Ratio: 4.928 Gaps: 1

Percent Similarity: 95.423 Percent Identity: 87.324

alignment_block:

US-09-557-262-2_COPY_21_307 x AA148186/rev

Align seq 1/1 to reverse of AA148186 from 1 to 924

2 ProAlaAlaProGlyProGlyLeuLysGlnGlnGlnLeuLeuArgLysVa 18

851 CTCTTCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 805

18 llelleleuProGluAspThrGlyGlyThrAlaSerAsnGlySert 35

804 GGGGAGGCTGGAGGATGAGTGGGATGAGTGGGATGAGTGGGATG 755

35 hrGlnGlnLeuProGlnThrIlellelleGlyValArgLysGlyGlyThr 51

754 CCCAGCAGTGGCGGAGACCATCATCGCGCTGCCCAAGGCGGCGC 705

52 ArgAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaAlaAla 68

704 CCGGCACTCTTGGAGATGTTAGGCTGACAGAGAGAGAGAGAGAG 655

68 uAsnGluValHisPhePheAspTrpGluGluHisTyrSerClnGlyLeu 85

654 GAAAGAGGCTGACCTTCTTCGAGCTGGAGAGAGATTAAGGACAG 605

85 lyTrpTrpLeuThrGlnMetProPheSerSerProHisGlnLeuThrVal 101

604 GCTGGTACCTCAGCCAGATGCGCTCTCTCTGCGCCACACACAC 555

102 GluLysThrProAlaTyrPheThrSerProLysValProGluArgIleHi 118

554 GAGAGACCCCGCGGTATTTACGCTGCCCAAGTGGCTGAGCGAGCT 505

118 sSerMetAsnProThrIleArgLeuLeuLeuLeuArgAspProSerG 135

504 CAGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 455

135 luArgValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLys 151

454 AGCGGCTGCTATCTGACCTACACCCAAAGTGTCTACACACATG 405

152 HisLysProTyrProProIleGluAspLeuLeuMetArgAspGlyArgLe 168

404 CAAG 355

168 uAsnLeuAspTyrLysAlaLeuAsnArgSerIleTyrHisAlaHisMetL 185

354 CAATATGCAATACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 305

185 euAsnTrpLeuArgPhePheProGluHisIlellelleValAspGly 201

304 AGAACTGGCTGGCTTTTTCGGCTGGCTGGCTGGCTGGCTGGCT 255

202 AspArgLeuLeuArgAspPheProGlnLeuLysValGluArgPhe 218

254 GAG 205

218 cLeuLysLeuSerProGlnIlellelleAsnAlaSerAsnPheTyrHisAsnLys 235

PD 09-AUG-2001.
 XX 30-JAN-2001: 2001WO-US00669.
 XX 04-FEB-2000: 2000US-0180312.
 XX 26-MAY-2000: 2000US-0207456.
 XX 30-JUN-2000: 2000US-0608408.
 XX 03-AUG-2000: 2000US-0632366.
 XX 21-SEP-2000: 2000US-0234687.
 XX 27-SEP-2000: 2000US-0236359.
 XX 04-OCT-2000: 2000US-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR.
 XX WPI: 2001-488447/52.
 XX Human genome-derived single exon nucleic acid probes useful for
 XX analyzing gene expression in human fetal liver -
 XX Claim 1: SEQ ID NO 3682: 638pp + sequence listing; English.
 XX The invention relates to a single exon nucleic acid probe for
 XX measuring human gene expression in a sample derived from human foetal
 XX liver. The single exon nucleic acid probes may be used for predicting,
 XX measuring and displaying gene expression in samples derived from human
 XX fetal liver. The present sequence is a single exon nucleic acid
 XX probe of the invention.
 XX Note: The sequence data for this patent did not form part of the
 XX printed specification, but was obtained in electronic format directly
 XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 452 BP: 86 A: 114 C: 155 G: 97 T: 0 other;

alignment_scores:
 Quality: 748.00 Length: 150
 Ratio: 5.123 Gaps: 0
 Percent Similarity: 97.333 Percent Identity: 89.333

alignment_block:

US-09-557-262-2_COPY_21_307 x ABA55377/rev ..

Align seq 1/1 to reverse of: ABA55377 from: 1 to: 452

69 AsnGluValHisPheAspTrpGluHisTyrSerGlnIleLeu 85
 451 AACGAGGTCACCTTCTCCACTGGGAGGAGCATTACAGCCACGCTTGGG 402
 85 YTrpTyrLeuThrGlnMetProPheSerSerProHisGlnLeuThrValG 102
 401 CTGGTAACTGAGCTAATGAGGCTTCTCTGTGATCAACAGCTACAGTGG 452
 102 IuLysThrProAlaTyrPheThrSerProLysValProGluAArgHis 118
 351 AGAAGACCCGCGCTATTTCATGAGCAAAATGCTTGAGAGAGTATA 402
 119 SerMetAsnProThrIleArgLeuLeuLeuLeuLeuArgAspProSerG 135
 301 ACATGAACCCCTCCATCCGCTGCTGCTCATCTCGGAGACCCGTCGGA 252
 135 uArqValLeuSerAspTyrThrGlnValLeuTyrAsnHisGlnLysH 152
 251 GGGATATATATGAGTATATATGAGTATATATATATATATATATATAT 202
 152 IuLysProTyrProProIleGluAspLeuMetArgAspGlyArgLeu 168
 201 ACAAGCCCTACCGCTATATGAGAGTCTCTGTGAGGATGAGATC 152
 169 AsnLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisAlaHisMetLe 185
 151 AATGTGGACTACAGGCGCTCAAGCGGAGGCTTACCAAGCTGCACATGCA 102

185 uAsnTrpLeuArgPheProLeuGlyHisIleHisIleValAspGlyA 202
 101 GAACCTGATGCTTTTTCCTGGTGGTACATCAATATCTGATGAGG 52
 202 sPArgLeuIleArgAspProPheTrpGluIleGlnLysValGluArgPhe 218
 51 ACCGCTCATACAGGAGCCCTTCCCTGAGATCCAAAAGGTCCAGAGTTC 2
 seq_name /SIN55/q3data/qoneseq/geneseq-emb1/NA2001A.DAT ABA25103

seq_documentation_block:

ID ABA25103 standard; DNA; 452 BP.

XX AC ABA25103;

XX 23-JAN-2002 (first entry)

XX Probe #3569 for gene expression analysis in human heart cell sample.

XX Human; gene expression; heart; microarray; vascular system; probe;

XX cardiovascular disease; hypertension; cardiac arrhythmia;

XX congenital heart disease; ss.

XX Homo sapiens.

XX WO200157274-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00666.

XX 04-FEB-2000; 2000US-0180312

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000US-0024263

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR.

XX WPI: 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX hearts -

XX Claim 1: SEQ ID NO 3569; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for

XX measuring human gene expression in a sample derived from human heart. The

XX present sequence is one such probe. The probes may be used for

XX predicting, measuring and displaying gene expression in samples derived

XX from the human heart via microarrays. By measuring gene expression, the

XX probes are useful for predicting, diagnosing, grading, staging,

XX monitoring and prognosing diseases of the human heart and vascular system

XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and

XX congenital heart disease.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 452 BP: 86 A: 114 C: 155 G: 97 T: 0 other;

alignment_scores:

Quality: 748.00 Length: 150
 Ratio: 5.123 Gaps: 0

Percent Similarity: 97.333 Percent Identity: 89.333

alignment_block:

DT 06-NOV-2001 (first entry)
 XX Human bone marrow expressed single exon probe SEQ ID NO: 3636.
 DE Human: bone marrow expressed exon; gene expression analysis; probe:
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
 KW Homo sapiens.
 OS
 XX WO200157276-A2.
 PN 09-AUG-2001.
 PD
 XX
 XX 30-JAN-2001: 2001WO-US00668.
 XX 04-FEB-2000: 2000US-0180312.
 PR 26-MAY-2000: 2000US-0207456.
 PR 30-JUN-2000: 2000US-0608408.
 PR 03-AUG-2000: 2000US-0632366.
 PR 21-SEP-2000: 2000US-0234687.
 PR 27-SEP-2000: 2000US-0236359.
 PR 04-OCT-2000: 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-488900/53.
 DR
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 XX Example 4; SEQ ID NO: 3636; 658pp + Sequence Listing; English.
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention.
 XX
 XX Sequence 452 BP; 86 A; 114 C; 155 G; 97 T; 0 other;
 SQ
 alignment_scores:
 Quality: 748.00 Length: 150
 Ratio: 5.123 Gaps: 0
 Percent Similarity: 97.333 Percent Identity: 89.333
 alignment_block:
 US-09-557-262-2_COPY_21_307 x AAK29079/rev ..
 Align seg 1/1 to reverse of: AAK29079 from: 1 to: 452
 69 AsnGluValHisPheAspTropGluHisTyrSerGlnGlyLeuG1 85
 |||||
 451 AACGAGTCCACTCTTCGACTGGCAGCATTACAGCCACGGCTTGGG 402
 85 yTrpTyrLeuThrGlnMetProPheSerSerProHisGlnLeuThrValG 102
 |||||
 401 CTGGTACCTCAGCCAGATGCCCTTCTCTGGCCACACACGCTCACAGTGG 352
 102 LuLyThrProAlaTyrPheThrSerProLysValProGluAArgHis 118
 |||||
 351 AGAAGACCCCGCGTATTTTCAGTCGCCCAAGTGCCTGAGCGAGTCTAC 302
 119 SerMetAsnProThrIleArgLeuLeuLeuLeuLeuLeuLeuLeuLeu 135
 |||||
 301 ACATGAACCCGCTCATCCGCTGCTCTATCTCTGGAGACCCGTCGGA 252
 135 uArgValLeuSerAspTyrThrGlnValLeuTyrAsnHisGlnLysH 152
 |||||
 251 GCGCGTGTATCTGACTACACCCCAAGTGTCTACACACCATGCAGNAGC 202

152 IsLysProTyrProIleGluAspLeuLeuMetArgAspGlyArgLeu 168
 |||||
 201 ACAAGCCCTACCCGTCATCGAGGAGTTCTTGGTCCGCGATGCGAGGCTC 152
 169 AsnLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisAlaHisMetLe 185
 |||||
 151 AATGTGGACTACAAGGCCCTCAGCCGAGGCTCTACCCAGCTGCACATGCA 102
 185 uAscTrpLeuArgPhePheProLeuGlyHisIleHisIleValAspGlyA 202
 |||||
 101 GAATGGCTGGGCTTTTTCGGGTGGCGCAATCATCATTTGTAAGCGGG 52
 202 spArgLeuIleArgAspProPheProGluIleGlnLysValGluArgPhe 218
 |||||
 51 ACCGCTCATCAGGAGCCCTTCCCTCAGATCCAAAGGTCGAGAGGTTTC 2
 seq_name: /SIDS5/qcgdata/geneseq/geneseq.emb./NA2001A.DAT.AAI35025
 seq_documentation_block:
 ID AAI35025 standard, DNA, 452 BP.
 XX
 AC AAI35025;
 XX
 DT 17-OCT-2001 (first entry)
 XX
 DE Probe #3711 used to measure gene expression in human placenta sample.
 XX
 KW Probe: microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200157272-A2.
 PN
 XX 09-AUG-2001.
 PD
 XX 30-JAN-2001: 2001WO-US00663.
 XX 04-FEB-2000: 2000US-0180312.
 PR 26-MAY-2000: 2000US-0207456.
 PR 30-JUN-2000: 2000US-0608408.
 PR 03-AUG-2000: 2000US-0632366.
 PR 21-SEP-2000: 2000US-0234687.
 PR 27-SEP-2000: 2000US-0236359.
 PR 04-OCT-2000: 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-488997/53.
 DR
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 XX Claim 25; SEQ ID No 3711; 654pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENPs).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 XX
 SQ Sequence 452 BP; 86 A; 114 C; 155 G; 97 T; 0 other;

alignment_scores:
 Quality: 748.00 Length: 150
 Ratio: 5.123 Gaps: 0
 Percent Similarity: 97.333 Percent Identity: 89.333
 alignment_block:

US-09-557-262-2_COPY_21_307 x AA145025/rev ..

Align seq 1/1 to reverse of. AA135025 from. 1 to: 452

```
69 AsnGluValHisPhePheAspTrpGluGluHisTyrSerGlnGlyLeuG1 85
|||||
451 AACGAGGTCCACTTCTTCGA-TGGAGAGCATTAAGGACAGGCTTGGG 402
85 yTrpTyrLeuThrGlnMetProPheSerSerProHisGlnLeuThrValG 102
|||||
401 CTGGTAACTAGGCTAGATGGCTTCTCTGGCCACACAGCAGCTCCACAGTGG 352
102 LuLysThrProAlaTyrPheThrSerProLysValProGluArgIleHis 118
|||||
351 AGAAGACGCGCGCGGTATTTCACGTGGCGCAAGTGGCTGAGCGAGTCTAC 302
119 SerMetAsnProThrIleArgLeuLeuLeuArgAspProSerG1 135
|||||
301 AACATGAAACCGGTCATCCGGCTGCTGCTCATCTGCGAGACCCGTGGA 252
135 uArgValIleuSerAspTyrThrGluValIleuTyrAsnHisLeuGlnIysH 152
|||||
251 CCGCGTCTCTATCTGACTACACCCCAAGTGTCTACAAACCACATCCAGAGC 202
152 iLysProTyrProProIleGluAspLeuLeuMetArgAspGlyArgLeu 168
|||||
201 ACAAGCCCTACCCGTCCTCATCGAGGAGTCTCTGTCGCGCATGCCAGGCTC 152
169 AsnLeuAspTyrIlysAlaLeuAsnArgSerLeuTyrHisAlaHisMetLe 185
|||||
151 AATGTGATTAAGAGCTCTCAAGCGTAAATCTATACAGGTGCATGCA 102
185 uAsnTrpLeuArgPhePheProLeuGlyHisIleHisIleValAspGlyA 202
|||||
101 GAATGGCTGGCTTTTTCGGGTGGCGCACATCCACATTGTGGACGGCG 52
202 spArgLeuIleArgAspProPheProGluIleGlnLysValGluArgPhe 218
|||||
51 ACCGCTCATACAGGAGCCGCTTCCCTCAGATCCAAAAGGTGAGAGGTTTC 2
```

OM 01: US-09-557-262-2_copy_21_307.tac GenEmbl * out_format . pfs
 Date: Jul 17, 2002 5:47 AM
 About: Results were produced by the GenCore software, version 4.5.
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framer.p2n model -DPV=xlp
 -O=/gen2_1/5SPH01/S00557-262-2/rnaseq/020521_161304_2557/app_query_ras*1_1558
 -DB=GenEmbl -QMT=fastap -SUF=fix.p2n.rge -GAPO=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LACAP=10.000 -LAPEXT=0.000
 -GAPUP=4.500 -GAPEXT=0.050 -XAPAP=10.000 -XAPEXT=0.500
 -XAPUP=6.000 -XAPEXT=7.000 -XAPAP=10.000 -XAPEXT=0.500
 -DELIN=6.000 -DELINEXT=7.000 -START=1 -MATRIX=blomsum62
 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=20 -THR_SCORE=pcr
 -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
 -NORM_ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=unlimited
 -USER=US09557262 -CGENL 1.11342 -NFEU=6 -ICPU=3 -LONGING
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-557-262-2_COPY_21_307
 Query length: 287
 Database: GenEmbl.*
 Database sequences: 1707656
 Database length: 187333701
 Search time (sec): 7102.980000

score_list:

score_list:	Strd Orig	zScore	EScore	Len	Documentation
gb_ro:BC009133	+ 1544.00	2684.05	3.90-141	1657	BC009133 Mus musculus, heparan
gb_ro:AF019385	+ 1544.00	2684.08	3.90-141	1685	AF019385 Mus musculus heparan
gb_hlg:AC084416.3	+ 1544.00	2641.03	9.96-139	110000	Continuation (4 of 6) of AC084416.3
gb_hlg:AC098713	+ 1544.00	2633.92	2.40-138	9261	AC098713 Mus musculus chromo
gb_hlg:AF177430	+ 1514.00	2637.35	1.50-138	9466	AF177430 Rattus norvegicus 3-0
gb_pr:AF019386	+ 1345.50	2338.74	6.90-122	1305	AF019386 Homo sapiens heparan
gb_pr:AC073991	+ 1345.50	2289.43	3.70-119	15938	AC073991 Homo sapiens RAC ch
gb_pr:AC066240	+ 1338.50	2276.72	1.90-118	165998	AC066240 Homo sapiens chromo
gb_pat:AX3137978	+ 745.50	1282.50	4.40-63	2115	AX3137978 Sequence 18 from Pat
gb_pr:AX3135498	+ 745.50	1237.89	1.40-60	165549	AX3135498 Human DNA sequen
gb_om:CF3A388534	+ 685.00	1191.01	5.60-58	519	CF3A388534 Canis familiaris mRNA
gb_pr:AF105374	+ 592.00	1014.41	3.90-48	1908	AF105374 Homo sapiens heparan
gb_pr:AF105375	+ 592.00	1014.41	3.90-48	1908	AF105375 Homo sapiens heparan
gb_ro:AF168992	+ 582.00	948.60	2.90-47	1666	AF168992 Mus musculus b. glycos
gb_pr:AF105376	+ 582.00	948.60	2.90-47	1666	AF105376 Homo sapiens heparan
gb_pr:AF105377	+ 582.00	948.60	2.90-47	1666	AF105377 Homo sapiens heparan
gb_pr:AF105378	+ 582.00	948.60	2.90-47	1666	AF105378 Homo sapiens heparan
gb_pr:AC093373	+ 563.00	959.96	3.70-45	2546	AC093373 Homo sapiens CDNA FL
gb_hlg:AF105378	+ 540.00	922.12	5.30-43	2311	AF105378 Homo sapiens heparan
gb_hlg:AC017305	+ 450.50	726.26	1.30-32	92745	AC017305 Prosopithecus m. laticus
gb_in:AC008337	+ 450.50	720.81	8.70-32	16480	AC008337 Prosopithecus m. laticus
gb_in:AE003511	+ 450.50	713.57	2.20-31	443007	AE003511 Prosopithecus m. laticus
gb_hlg:AC027102	+ 437.00	674.28	2.70-29	168472	AC027102 Homo sapiens chromo
gb_pr:AC093373	+ 437.00	674.28	2.70-29	168472	AC093373 Homo sapiens chromo
gb_pr:AC098715	+ 425.00	672.19	4.40-29	249269	AC098715 Mus musculus chromo
gb_hlg:AC093359	+ 424.00	671.79	4.70-29	218502	AC093359 Mus musculus clone
gb_pr:AC093511	+ 423.00	671.51	4.90-29	189244	AC093511 Homo sapiens chromo
gb_hlg:AL603889	+ 412.00	646.85	1.10-27	320066	AL603889 Mus musculus chromo
gb_hlg:AC021445	+ 410.50	648.86	8.90-28	203621	AC021445 Mus musculus clone
gb_in:GELH15803	+ 409.50	674.00	2.70-29	12172	AC098996 Genethal 1.3 clone
gb_pr:AF098987	+ 404.50	668.69	7.00-29	24810	AF098987 Canis familiaris clone
gb_pr:AF006640	+ 404.50	668.69	7.00-29	24810	AF006640 Homo sapiens lopl.3
gb_pr:HS43946	+ 402.00	650.34	7.30-28	41255	AF031723 Human DNA sequen
gb_pr:AC005353	+ 399.00	658.44	4.20-27	75108	AC005353 Homo sapiens chromo
gb_hlg:AC012140	+ 399.00	626.53	1.60-26	251947	AC012140 Homo sapiens chromo
gb_pr:AC005375	+ 394.50	624.04	2.10-26	149030	AC005375 Homo sapiens chromo
gb_pr:AC005234	+ 392.00	618.50	3.30-26	166647	AC005234 Homo sapiens chromo
gb_in:AF175689	+ 348.00	577.48	8.40-24	4947	AF175689 Prosopithecus m. laticus
gb_pr:AF015649	+ 343.50	573.57	1.40-23	3359	AF015649 Homo sapiens NDST4 mR
gb_ro:AB036844	+ 337.00	565.17	4.10-23	2732	AC036838 Mus musculus NDST4 mR
gb_pr:AF076605	+ 336.50	562.54	5.70-23	2978	AF076605 Homo sapiens heparan

ql_r:AF074524 396.50 551.84 6.50-23 2188 AF074524 Homo sapiens hepar
 ql_r:ELP5461 396.50 551.84 6.50-23 2204 ELP5461 Homo sapiens elp
 ql_r:HS076601 396.50 551.84 6.50-23 2221 HS076601 Homo sapiens heparan
 ql_r:G027044 396.50 551.84 6.50-23 2231 G027044 SMO2.31617 Haman Hom
 ql_r:AF0221095 396.50 551.84 6.50-23 2783 AF0221095 Mus musculus hepar

seq_name: qb_ro:BC009133

seq_documentation_block:

seq_id: BC009133

DEFINITION

Mus musculus, heparan sulfate (glucosamine) 3'-O-sulfotransferase 1,

clone W3-11450 IMAGE:415049, mRNA, complete cds.

ACCESSION

BC009133.1 GI:14318652

VERSION

MGC.

KEYWORDS

house mouse.

SOURCE

Mus musculus.

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Scurionathia; Muridae; Mus;

Strasbourg, R.

Direct Submission

TITLE

Submitted (05 JUN 2001) National Institutes of Health, Mammalian

Gene Collection, (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

PMAPK

Contact: MGC help desk

Email: mgc@nci.nih.gov

Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I M A G E Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome

Sequencing Center

Center code: BCM-HGSC

Web site: http://www.bcm.tmc.edu/cdna/

Contact: vialla@bcm.tmc.edu

Vialla, V.K., Luna, P.A., Hale, S.M., Bullyk, S., Lu, X., Garcia,

A.M., Hellmeyer, M., Telford, H., Redfearn, A., Buck, J., Yu, W.,

Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found

through the I M A G E Consortium/LLNL at: http://image.llnl.gov

Series: IPAK plater 16 Row, 6 Column: 17

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA 3' 5518970.

Location/Qualifiers

1. 1657

/organism="Mus musculus"

/db_xref="LocusID:15476"

/db_xref="taxon:10090"

/accession="M211450 IMAGE:3155049"

/tissue="type="Mammary tumor, H&E, 7d, MMIV 3rd model, 10

months old, gross tissue."

/product="3-O-sulfotransferase 1"

/protein="P1" "MAM3"

/db_xref="GI:14318652"

/feature="CD10B"

/note="Vector: pCMV-SPORT6"

284 1219

/feature="start=1"

/product="heparan sulfate (glucosamine)

3-O-sulfotransferase 1"

/protein="P1" "MAM3"

/db_xref="GI:14318652"

/feature="CD10B"

/note="Vector: pCMV-SPORT6"

284 1219

/feature="start=1"

/product="heparan sulfate (glucosamine)

3-O-sulfotransferase 1"

/protein="P1" "MAM3"

/db_xref="GI:14318652"

/feature="CD10B"

/note="Vector: pCMV-SPORT6"

284 1219

/feature="start=1"

/product="heparan sulfate (glucosamine)

3-O-sulfotransferase 1"

/protein="P1" "MAM3"

/db_xref="GI:14318652"

/feature="CD10B"

/note="Vector: pCMV-SPORT6"

284 1219

/feature="start=1"

/product="heparan sulfate (glucosamine)

3-O-sulfotransferase 1"

/protein="P1" "MAM3"

/db_xref="GI:14318652"

/feature="CD10B"

/note="Vector: pCMV-SPORT6"

284 1219

/feature="start=1"

/product="heparan sulfate (glucosamine)

3-O-sulfotransferase 1"

/protein="P1" "MAM3"

/db_xref="GI:14318652"

/feature="CD10B"

/note="Vector: pCMV-SPORT6"

284 1219

/feature="start=1"

/product="heparan sulfate (glucosamine)

3-O-sulfotransferase 1"

/protein="P1" "MAM3"

/db_xref="GI:14318652"

/feature="CD10B"

/note="Vector: pCMV-SPORT6"

284 1219

/feature="start=1"

/product="heparan sulfate (glucosamine)

3-O-sulfotransferase 1"

/protein="P1" "MAM3"

/db_xref="GI:14318652"

/feature="CD10B"

/note="Vector: pCMV-SPORT6"

284 1219

/feature="start=1"

/product="heparan sulfate (glucosamine)

3-O-sulfotransferase 1"

/protein="P1" "MAM3"

/db_xref="GI:14318652"

/feature="CD10B"

/note="Vector: pCMV-SPORT6"

284 1219

/feature="start=1"

/product="heparan sulfate (glucosamine)

3-O-sulfotransferase 1"

/protein="P1" "MAM3"

/db_xref="GI:14318652"

/feature="CD10B"

/note="Vector: pCMV-SPORT6"

284 1219

/feature="start=1"

/product="heparan sulfate (glucosamine)

3-O-sulfotransferase 1"

/protein="P1" "MAM3"

/db_xref="GI:14318652"

/feature="CD10B"

/note="Vector: pCMV-SPORT6"

284 1219

/feature="start=1"

/product="heparan sulfate (glucosamine)

3-O-sulfotransferase 1"

/protein="P1" "MAM3"

/db_xref="GI:14318652"

/feature="CD10B"

/note="Vector: pCMV-SPORT6"

284 1219

/feature="start=1"

/product="heparan sulfate (glucosamine)

3-O-sulfotransferase 1"

/protein="P1" "MAM3"

/db_xref="GI:14318652"

/feature="CD10B"

/note="Vector: pCMV-SPORT6"

284 1219

/feature="start=1"

/product="heparan sulfate (glucosamine)

3-O-sulfotransferase 1"

/protein="P1" "MAM3"

/db_xref="GI:14318652"

/feature="CD10B"

/note="Vector: pCMV-SPORT6"

284 1219

/feature="start=1"

/product="heparan sulfate (glucosamine)

3-O-sulfotransferase 1"

/protein="P1" "MAM3"

/db_xref="GI:14318652"

/feature="CD10B"

/note="Vector: pCMV-SPORT6"

284 1219

/feature="start=1"

/product="heparan sulfate (glucosamine)

3-O-sulfotransferase 1"

/protein="P1" "MAM3"

/db_xref="GI:14318652"

/feature="CD10B"


```

/product="3-O-sulfotransferase"
/protein_id="AAC09283.1"
/db_xref="gi:9457244"
/translation="MTLLLLGAVLLVAQPOLVPSHPAPGGLKQQLLRKVIILPED
TUGKATNTSGOGLPDTITTVRKGGTRALLFEMLSLDPDVAANEVIFPFWEHYSQ
GLGWYNTNPFSSPHQIVKETIPAYETSPKVKPERIHSNPTIRLLLIIRDFSERVLSQ
YGVLYNHIAQKKKPYPPIDILIMROGRINVDYKALNKSLSYAHMMINWIRFPPLVSHIH
VQDRPILRPDPFLEIKVERFKLSPIINASNPFYKTKGFCVLRUSCKDKCLHRSCKM
AUPQVDPKLLDLKHEVEREPNKKPKLVGRFDWH"
BASE COUNT      224  292  224  196
ORIGIN

```

```

alignment_scores:
  Quality: 1514.00      Length: 287
  Ratio: 5.294          Gaps: 0
  Percent Similarity: 99.652  Percent Identity: 97.909

```

```

alignment_block:
US-09-557-262-2_copy_21_307 x AF177430

```

```

Align seq 1/1 to: AF177430 from: 1 to: 936

```

```

1  HisproAlaAlaProGlyProGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 17
|||||
61  CACCGCGCTGCTCTGGCCCGGACCTCAACACACACACGCGCTTCTGAGGAA 110
|||||
17  sValIleIleLeuProGluAspThrGlyGlyGlyGlyThrAlaSerAsnGly 34
|||||
111  GCGCATATATGCGCGGAGCAACAAGAGAAAGGAGGAGGAGGAGGAGGAG 160
|||||
34  erThrGlnGlnLeuProGlnThrIleIleIleGlyValArgGlyGlyGly 50
|||||
161  GCAAGCAGAGAGTGGCAACAATATATATATTGATTTGATTTGATTTGAT 210
|||||
51  ThrArgAlaIleuLeuGluMetLeuSerLeuHisProAspValAlaAla 67
|||||
211  ACCGACGCGCTGCTGGAGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 260
|||||
67  aGluAsnGluValHisPhePheAspTTPGluGluHisTyrSerGlnGly 84
|||||
261  TCAAAACACGCTGCAATCTCTTCAGCTGGGAGGAGGAGGAGGAGGAGGAG 310
|||||
84  ouGlyTTPGlyThrGlnMetProPheSerSerProHisGlnIleThr 100
|||||
311  TGAGTGTGATCTACCAAGATGCGCTTCTCTGCTGCTGCTGCTGCTGCTG 360
|||||
101  ValGlyTyrThrProAlaTyrPheThrSerProGlyValProGluArg 117
|||||
361  GTAGAGAGACACCTGCTTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 410
|||||
117  ellisSerMetAsnProThrIleArgLeuLeuLeuLeuLeuLeuLeuLeu 134
|||||
411  CACAGCAGTGAACGCTACTATCGGCGGCTGCTGCTGCTGCTGCTGCTGCT 460
|||||
134  erGluArgValIleuSerAspTyrThrGlnValIleuTyrAsnHisIleu 150
|||||
461  CGGAGAGTGTGATATGAGATACACAGGAGGAGGAGGAGGAGGAGGAGGAG 510
|||||
151  LysHisGlyProTyrProTyrIleGluAspIleuMetArgAspGlyArg 167
|||||
511  AAGCAAAACCTTACCCACCCATTTGAGCAGCTGCTGCTGCTGCTGCTGCT 560
|||||
167  uLeuAsnLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisAlaHis 184
|||||
561  GCTTAACGTCGATACAGGCTCTCAACCCAGCCCTGCTACCATGCGACACA 610
|||||
184  etLeuAsnTyrPheArgPhePheProLeuGlyHisIleHisIleValAsp 200
|||||
611  TGCTGCAATGCTGATGTTTTTTTCTGCAAGGAGGAGGAGGAGGAGGAGG 560
|||||
201  GlyAspArgTyrLeuArgAspProPheThrGluThrGlyGlyValGlu 217
|||||

```

```

661  GCGGACCGCTTTCATCAGAGACCGCTTCCCTCAGATTCAGAAAGCGAAAG 710
|||||
217  oPheLeuLysLeuSerProGlnIleAsnAlaSerAsnPheTyrPheAsnL 234
|||||
711  GPDGCTAAGTGTCTCAAGATCAAGATCAAGATCAAGATCAAGATCAAGAT 760
|||||
234  ySPHRTSCPTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 250
|||||
761  AAACCAAGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 810
|||||
251  HisGluSerLysGlyArgAlaHisProGlnValArgProLysLeuLeuAs 267
|||||
811  CATGAGTCCAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 860
|||||
267  rLysLeuLysGlyArgPheHisGluProAsnLysGlyPhePheGlyLeu 284
|||||
861  TAACTGCAAGAAATATTTCTGAGAGCAATATAGAAATTTTTCAGGTGG 910
|||||
284  aIGlyArgThr 287
|||||
911  TGGGTCAGAAACA 921

seq_name: qb_pr:AF019386
seq_documentation_block:
LOCUS      AF019386                1305 bp    mRNA    linear    191 14 Nov 1997
DEFINITION Homo sapiens heparan sulfate 3-O-sulfotransferase-1 precursor
ACCESSION  AF019386
VERSION    AF019386.1    GI:2618972
KEYWORDS   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Gracilata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 1305)
AUTHORS   Shworak, N.W., Liu, J., Fritze, D.M.S., Schwartz, J.J., Zhang, L.,
            Logeart, D. and Rosenberg, R.D.
            Molecular cloning and expression of mouse and human cDNAs encoding
            heparan sulfate D-glucosaminyl 3-O sulfotransferase
            J. Biol. Chem. 272 (44): 28008-28019 (1997)
MEDLINE   98010647
REFERENCE  2 (bases 1 to 1305)
AUTHORS   Shworak, N.W., Liu, J., Fritze, D.M.S., Schwartz, J.J., Zhang, L.,
            Logeart, D. and Rosenberg, R.D.
            Direct Submission
            Submitted (14-AUG-1997) Biology, MIT, 31 Ames Street, Cambridge, MA
            02139, USA
FEATURES   Location/Qualifiers
            1..1305
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /note="Isolated from lamda Triplex Brain cDNA library
            from Clontech"
            1..1305
            /gene="3OST1"
            1..118
            /gene="3OST1"
            119..178
            /gene="3OST1"
            119..1042
            /gene="3OST1"
            /function="rate limiting enzyme for synthesis of
            antileucinant heparan"
            /note="heparan sulfate sulfotransferase; 3-OST-1;
            interluminal Golgi resident protein (retension mechanism
            unknown)"
            /codon_start=1
            /product="heparan sulfate 3-O-sulfotransferase-1
            precursor"
            /protein_id="AAC84388.1"
            /db_xref="gi:2618973"
            /translation="MAALLGCAVLLVAQPOLVPSHPAPGGLKQQLLRKAGTLLQDVRKIG

```


TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (08-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA

4 (bases 1 to 159838)
Waterston,R.H.
Direct Submission
Submitted (03-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA

5 (bases 1 to 159838)
Waterston,R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 26, 2001 this sequence version replaced gi.8480614

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/qsc/>
Contact: sapient@wustl.wustl.edu
Summary Statistics
Center project name: H_NA0512120

NOTE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats, all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John B. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/qsc>

SOURCE INFORMATION:
The RP11-11 human HAC library was made from the blood of one male donor, as described by Otagawa, K., Woon, P.Y., Zhao, R., Fung, C.E., Tateo, M., Calanese, J.J., and de Jond, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-669F3; the clone sequenced to the right is AC056230. Actual start of this clone is at base position 1 of RP11-512120; actual end is at base position 159838 of RP11-512120.

FEATURES
source
Location/Qualifiers
1..159838
organism="Homo sapiens"
db_xref="taxon:9606"
chromosome="4"
map="4"
clone="RP11-512120"
clone_lib="RP11-11"
2..775
repeat_region
266..317
repeat_region
776..792
repeat_region
rp1_family="MERL_type"
rp1_family="I-rich"
rp1_family="AT-rich"

repeat_region
845..2469
rp1_family="LI"
repeat_region
2470..2766
rp1_family="Alu"
repeat_region
2769..3291
rp1_family="MERL_type"
repeat_region
3321..3441
rp1_family="LI"
repeat_region
3628..3655
rp1_family="AT-rich"
repeat_region
4297..4409
rp1_family="MIR"
repeat_region
4815..5120
rp1_family="Alu"
repeat_region
5094..5114
rp1_family="(A)n"
repeat_region
5203..5546
rp1_family="L2"
repeat_region
8632..8739
rp1_family="GA-rich"
repeat_region
9549..9890
rp1_family="ERV1"
repeat_region
10107..10349
rp1_family="MIR"
repeat_region
11853..12152
rp1_family="MALR"
repeat_region
12209..12271
rp1_family="LI"
repeat_region
12305..12390
rp1_family="L2"
repeat_region
12605..12876
rp1_family="LI"
repeat_region
12877..13185
rp1_family="Alu"
misc_feature
13162..13185
/note="similar to Homo sapiens EST BF181502 (NM:0106181)"
repeat_region
13186..13221
rp1_family="LI"
repeat_region
13223..13488
rp1_family="LI"
repeat_region
13489..13750
rp1_family="Alu"
repeat_region
13751..14291
rp1_family="LI"
repeat_region
14332..14380
rp1_family="MERL_type"
repeat_region
15690..15714
rp1_family="(T)n"
repeat_region
15873..16172
rp1_family="Alu"
repeat_region
16173..16300
rp1_family="Alu"
repeat_region
16910..16942
rp1_family="AT-rich"
repeat_region
17130..17499
rp1_family="ERV1"
repeat_region
17303..17479
rp1_family="A-rich"
repeat_region
17553..17621
rp1_family="(CATATA)n"
repeat_region
17920..17970
rp1_family="GA-rich"
repeat_region
18178..18658
rp1_family="LI"
repeat_region
18644..18953
rp1_family="LI"
repeat_region
19026..19155
rp1_family="L2"
repeat_region
19061..19084
rp1_family="AT-rich"
repeat_region
19167..19441
rp1_family="LI"

```

repeat_region 19638..19749
/rpt_family="L1"
repeat_region 19850..20000
/rpt_family="L1"
repeat_region 20251..20404
/rpt_family="MER2_type"
repeat_region 21233..22285
/rpt_family="L1"
repeat_region 23307..23479
/rpt_family="MER1_type"
repeat_region 24526..24729
/rpt_family="(CTA)n"
repeat_region 24570..24849
/rpt_family="Alu"
repeat_region 25680..25700
/rpt_family="(TA)n"
repeat_region 26134..26688
/rpt_family="ERV1"
repeat_region 26944..27375
/rpt_family="MALR"
repeat_region 27885..28356
/rpt_family="L2"
repeat_region 28532..28641
/rpt_family="L2"
repeat_region 28723..28942
/rpt_family="MIR"
repeat_region 29079..29542
/rpt_family="L2"
repeat_region 29566..29888
/rpt_family="L2"
repeat_region 29899..30411
/rpt_family="ERV1"
repeat_region 30439..30539
/rpt_family="L2"

alignment_scores:
Quality: 1345.50 length: 296
Ratio: 4.929 Gaps: 1
Percent Similarity: 95.455 Percent Identity: 87.413

alignment_block:
US-09-557-262-2_copy_21_307 x AC073991/rev ..
Align seq 1/1 to reverse of: AC073991 from: 1 to: 159838

2 ProAlaLaProGlyProGlyLeuLysGlnGlnGlnLeuLysVal 18
||||| ||||| ||| ||||| ||||| ||||| |||||
155036 CCHCCCGCGCGCGGAG CTAAGCCAGCCAGGTTCTGCGGAAAGC 154940

18 lilelleLeuProGluAspThrGlyGluGlyThrAlaSerAsnGlySer 35
||||| ||||| ||||| ||||| ||||| ||||| |||||
154989 GAGGAGGCTGAGATGAGTGGATGGAGGAGGAGGAGGAGGAGGAG 154940

35 hrgGlnGlnProGlnThrIlelleGlyValArgLysGlyGlyThr 51
||||| ||||| ||||| ||||| ||||| ||||| |||||
154939 CCGACAGTGGCGCAGACCAATCATCGCGGCGGCGGCGGCGGCGG 154890

52 ArqAlaLeuLeuGluMetLeuSerLeuHisProAspValAlaAla 68
||||| ||||| ||||| ||||| ||||| ||||| |||||
154889 CCGCGACCTGCTGGAGATGCTGAGCTGCGGCGGCGGCGGCGGCGG 154840

68 uAsnGluValHisPheAspThrGluGluHisTyrSerGlnGlyLeu 85
||||| ||||| ||||| ||||| ||||| ||||| |||||
154839 GAGGAGGCTGAGATGAGTGGATGGAGGAGGAGGAGGAGGAGGAG 154790

85 lrrPtyrLeuThrGlnMetProPheSerSerProHisGlnGlnThrVal 101
||||| ||||| ||||| ||||| ||||| ||||| |||||
154789 GTGCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 154740

102 GluLysThrProAlaTyrPheThrSerProLysValProGluArqlle 118
||||| ||||| ||||| ||||| ||||| ||||| |||||
154739 CAGAGACCGCGCGGATGATTCAGCGGCGGCGGCGGCGGCGGCGG 154690

```

```

118 sserMetAsnProThrIleArgLeuLeuLeuLeuLeuLeuLeuLeu 135
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154889 CAGATATTAAGGCTGATGAGTGGATGGAGGAGGAGGAGGAGGAG 154840

135 lArqValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeuGlnLys 151
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154639 AGCGCGTCTATCTGACTACACCGCAAGTGTCTACACCCACATCC 154590

152 HisLeuPheTyrProPheIleGluAspLeuLeuMetArgAsnGly 168
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154589 CACAAAGGCTATAGGCTGATGAGTGGATGGAGGAGGAGGAGGAG 154540

168 uAsnLeuAspTyrLysAlaLeuAsnArgSerLeuTyrHisAlaHisMet 185
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154539 CAATTTGAGTAAAGAGGCTGAAAGGCTGAAAGGCTGAAAGGCTG 154490

185 euAsnThrLeuArgPhePheProLeuGlyHisIleHisIleValAsp 201
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154489 ACAAAGGCTGATGAGTGGATGGAGGAGGAGGAGGAGGAGGAGG 154440

202 AspArgLeuIleArgAspProPheProGluIleGlnLysValGlnArg 218
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154439 GACCGGCTCATCAAGAGAGGCTGAGGAGGAGGAGGAGGAGGAGG 154390

218 eLeuLysLeuSerProGlnIleAspAlaSerAspPheTyrThrAsn 235
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154389 CCAAAAGGCTGATGAGTGGATGGAGGAGGAGGAGGAGGAGGAG 154340

235 hrLysGlyPheTyrCysLeuArgAspSerGlyLysAspArgCysLeu 251
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154339 CCAAGGCTGATGAGTGGATGGAGGAGGAGGAGGAGGAGGAGGAG 154290

252 GluSerLysGlyArgAlaHisProGlnValAspProLysLeuAsp 268
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154289 GAGTCCAAAGGCTGATGAGTGGATGGAGGAGGAGGAGGAGGAGG 154240

268 sLeuHisGlnTyrThrHisGlnIleProAsnLysLysPheLeuLeu 285
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
154239 ACTGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 154190

285 lYArgThr 287
|||||
154189 GCAGAAC 154182

seq_name: gb.pr.AC006230
seq_documentation_block:
LOCUS Arhuh249
DEFINITION Homo sapiens chromosome 4 clone C0287314 map 4p16, complete
sequence.
AC006230.11 GI:4966389
VERSION Arhuh249
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 166998)
AUTHORS Stone,N.E., Schmutz,J.J., Sheng,J., Pennacchio,L.A., Cox,D.P. and
Myers,R.M.
JOURNAL Direct Submission
TITLE Unpublished
REFERENCE 2 (bases 1 to 166998)
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.P. and Myers,R.M.
JOURNAL Direct Submission
TITLE Submitted (18-JUL-1998) Department of Genetics, Stanford Human
Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
REFERENCE 3 (bases 1 to 166998)
AUTHORS Stone,N.E., Schmutz,J.J., Sheng,J., Pennacchio,L.A., Cox,D.P. and
Myers,R.M.
JOURNAL Direct Submission
TITLE Submitted (02-MAR-1998) Department of Genetics, Stanford Human
Genome Center, 4005 Miranda Avenue, Palo Alto, CA 94304, USA

```



```

/Note-"15 copies 6 mer tatata 72% conserved"
7185..7230
/Note-"18 copies 2 mer ta 77% conserved"
8044..8192
/Note-"MIR repeat: matches 22..237 of consensus"
8200..8415
/Note-"MER6JB repeat: matches 209..436 of consensus"
8416..8708
/Note-"AluSq repeat: matches 1..293 of consensus"
8709..8954
/Note-"MER6JB repeat: matches 1..209 of consensus"
8983..9031
/Note-"MIR repeat: matches 193..241 of consensus"
9130..9173
/Note-"MIR repeat: matches 76..119 of consensus"
9884..10051
/Note-"LIM3 repeat: matches 6740..6807 of consensus"
10112..10172
/Note-"M11A1 repeat: matches 1..365 of consensus"
10338..10734
/Note-"MER6BA repeat: matches 22..224 of consensus"
10747..10856
/Note-"L1M4 repeat: matches 3607..3737 of consensus"
10885..11176
/Note-"L2 repeat: matches 2471..2627 of consensus"
11273..11438
/Note-"AluIo repeat: matches 144..207 of consensus"
11439..11742
/Note-"AluSc repeat: matches 1..304 of consensus"
11743..11788
/Note-"AluIo repeat: matches 97..144 of consensus"
11885..11967
/Note-"L2 repeat: matches 2637..2718 of consensus"
13808..13919
/Note-"MIR repeat: matches 35..147 of consensus"
14580..14874
/Note-"AluSc repeat: matches 1..295 of consensus"
15859..15940
/Note-"match: STS: EM-60413"
15712..15779
/Note-"34 copies 2 mer ca 79% conserved"
15716..15775
/Note-"10 copies 6 mer caaca 81% conserved"
17139..17258
/Note-"FLAM-A repeat: matches 1..122 of consensus"
18171..18700
/Note-"MLTID repeat: matches 2..505 of consensus"
18863..18939
/Note-"AluIo/FLAM repeat: matches 43..124 of consensus"
18940..19217
/Note-"AluIo repeat: matches 1..304 of consensus"
19286..19447
/Note-"MLTII repeat: matches 118..264 of consensus"
19558..19694
/Note-"MLTII repeat: matches 345..483 of consensus"
20090..21304
/Note-"Tiger3b repeat: matches 1..1230 of consensus"
22028..22075
/Note-"L2 repeat: matches 2458..2508 of consensus"
23080..23346
/Note-"AluSq repeat: matches 41..308 of consensus"
29225..29588
/Note-"THEIC repeat: matches 1..371 of consensus"
29765..30063
/Note-"AluSq repeat: matches 1..299 of consensus"
31420..31708
/Note-"AluSq repeat: matches 1..292 of consensus"
32985..33283
/Note-"AluSq repeat: matches 1..301 of consensus"
34330..34786
/Note-"L2 repeat: matches 2248..2703 of consensus"
44858..45006
/Note-"MIR repeat: matches 79..248 of consensus"

```

```

repeat_region 35193..35341
/Note-"MIR repeat: matches 48..194 of consensus"
36032..36094
/Note-"MER5A repeat: matches 88..152 of consensus"
36125..36410
/Note-"AluSq repeat: matches 1..366 of consensus"
37642..37713
/Note-"MIR repeat: matches 106..180 of consensus"
38972..39030
/Note-"MIR repeat: matches 148..199 of consensus"
39074..39210
/Note-"MIR repeat: matches 31..178 of consensus"
39269..39573
/Note-"AluSq repeat: matches 1..303 of consensus"
39700..39823
/Note-"L1B2 repeat: matches 5930..6152 of consensus"
40412..40941
/Note-"L2 repeat: matches 1813..2407 of consensus"
40961..41060
/Note-"MIR repeat: matches 131..233 of consensus"
41058..41172
/Note-"L2 repeat: matches 2406..2710 of consensus"
42145..42381
/Note-"MER63 repeat: matches 745..998 of consensus"
42430..44048
/Note-"L1PA7 repeat: matches 4526..6144 of consensus"
44072..44167
/Note-"MER63 repeat: matches 2..734 of consensus"
44787..44901
/Note-"L2 repeat: matches 2609..2747 of consensus"
45261..45368
/Note-"MER45 repeat: matches 73..1027 of consensus"
47949..48247
/Note-"AluSq repeat: matches 2..308 of consensus"
48321..48466
/Note-"MIR repeat: matches 37..191 of consensus"
48586..49514
/Note-"TIGGR1 repeat: matches 1489..2417 of consensus"
49515..49828
/Note-"AluSp repeat: matches 1..608 of consensus"
49829..50088
/Note-"TIGGR1 repeat: matches 1230..1489 of consensus"
50148..50693
/Note-"L1PA16 repeat: matches 5603..6157 of consensus"
50716..51171
/Note-"TIGGR1 repeat: matches 4..475 of consensus"
51605..51816
/Note-"L1M88 repeat: matches 5956..6171 of consensus"
52358..52403
/Note-"MER67A repeat: matches 449..543 of consensus"
52411..52488
/Note-"L1TR10A repeat: matches 13..98 of consensus"

alignment_scores
Quality: 745.50 Length: 290
Ratio: 3.404 Gaps: 3
Percent similarity: 75.517 Percent identity: 48.966

alignment_block:
US-09-557-262-2_COPY_21_307 x AL355498/rev
Align seq 1/1 to reverse of: AL355498 from: 1 to: 165549
1 HisProAlaAlaProGlyProGlyLeuLysGlnIleGluLeuLeuArgArg 17
|||||
27845 CACCAGTTCGGGAAGCGCAAGCGCTCCCAAGAGCAAGTTCGCGTCCATCA 27796
17 sValIleIleLeuLeuGluAspThrGlyGlyIleAlaSerAsnGlyS 34
27795 CCGGGTC
27789
34 GGGGAGGlnGlnGlnProGlnThrIleIleIleGlyValArgLysGlyC 50
|||||

```



```

100 thrValGluGlyThrProAlaTyrPheThrSerProIlysValProGluAr 116
116 qlleHisSerMetAsnProThrIleArgLeuLeuLeuLeuLeuLeuLeuLeu 133
133 ::::|||||
321 ACGGTGAGAAAGAGGAGGATTAATTAAGGAGGAGGAGGAGGAGGAGGAGGAG 370
116 qlleHisSerMetAsnProThrIleArgLeuLeuLeuLeuLeuLeuLeuLeu 133
133 ::::|||||
371 GGTGCACAGATGACAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
133 roSerGluArgValLeuSerAspTyrThrGlnValLeuTyrAsnHisLeu 149
149 ::::|||||
421 GGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 470
150 GlnHisHisLysProTyrProProIleGluAspIleuLeuMetArgAsp 165
165 ::::|||||
471 CAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 518

seq_name: qb_pr:AF105374

seq_documentation_block:
LOCUS      AF105374                1468 bp    mRNA    linear    PRI: 11-JUN-2001
DEFINITION Homo sapiens heparan sulfate D-glucosaminyl 3-O-sulfotransferase-2
(3OST2) mRNA, 3OST2-1 allele, complete cds.
ACCESSION  AF105374
VERSION    AF105374.1  GI:4835718
KEYWORDS
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi,
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 1968)
            Shworak,N.W., Liu,J., Petros,I.M., Zhang,L., Kobayashi,M.,
            Copeland,N.G., Jenkins,N.A. and Rosenberg,R.D.
            Multiple isoforms of heparan sulfate D-glucosaminyl
            3-O-sulfotransferase. Isolation, characterization, and expression
            of human cDNAs and identification of distinct genomic loci
            J. Biol. Chem. 274 (8), 5170-5184 (1999)
            99143187
REFERENCE  2 (bases 1 to 1968)
            Shworak,N.W., Liu,J. and Rosenberg,R.D.
            Direct Submission
            Submitted (12-NOV-1998) Biology, Massachusetts Institute of
            Technology, 31 Ames St, Cambridge, MA 02139, USA
FEATURES             location/qualifiers
     source          1..1968
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="16"
                    /map="16p12"
                    /tissue_type="brain"
                    /note="isolated by hybridization screening using IMAGE
                    Consortium clone c-20d10 found in GenBank Accession Number
                    F07258"
     gene            1..1968
                    /gene="3OST2"
                    /allele="3OST2-1"
                    <1..72
                    /gene="3OST2"
                    73..1176
                    /gene="3OST2"
                    /acc_number="2.9.2.23"
                    /note="3-OST-2"
                    /codon_start=1
                    /product="heparan sulfate D-glucosaminyl
                    3-O-sulfotransferase-2"
                    /protein_id="RAD30206.1"
                    /db_xref="GI:4835719"
     5'UTR          1102..1110
                    /gene="3OST2"
                    /note="putative N-glycosylation site"
     CDS             1177..1950
                    /gene="3OST2"
                    /note="putative cystine bridged peptide loop; unclassified
                    site"
                    1249
                    /gene="3OST2"
                    /note="3OST2 2 allele"

```



```

131 ArgAspProSerGluArgValLeuSerAspTyrThrGlnValLeuTyrAs 147
|||||
1401012 CAGAGCAAGGATGAAGAGGACATCTGGAGTAAAGAGAAAG 1053
|||||
147 nHisLeuGlnLysHisLysProTyrProProTleGluAspLeuLeuMetA 164
|||||
10054 CAGTTCACAGCGCGCGCGACACCGTAACTTTGAGAGAGCTTGAAGCTTCA 1099
|||||
164 rgAsp.....GlyArgLeuAsnLeuAspTyrLysAlaLeuAsnArg 177
|||||
111100 AAAACAGGACAGCGCGCGCTCAATGACACGGTGGAGGCGCCATCCAGATC 1149
|||||
178 SerLeuTyrHisAlaHisMetLeuAsnTTPLeuArgPhePheProLeuG 194
|||||
111150 GGAATCATAGCAAGAAATCTGAGAAATGAGTATGAAATATGGAATAGG 1199
|||||
194 yHisTleHisTleValAspGlyAspAlaGluLeuAlaArgAspProPheProG 211
|||||
1200 CCAGATGCTCTTCGTGAGCGCGAGCGGGCTCAACAGCGACCTGGCGCGGG 1249
|||||
211 luleGlnLysValGluArgPheLeuLysLeuSerProGlnTleAsnAla 227
|||||
1250 AGCTGGCGCGCGGTCGAAGACATCTCTGGGCGCTCAACATGATCATCATGAGAC 1299
|||||
228 SerAsnHisTyrPheAsnLysThrLysGlyPheTyrCysLeuArg.... 242
|||||
1300 AAGAAATTTACTTTAAATAAAGAAAGAAAGAGTTGGCTGGTAAAGAGAG 1349
|||||
243 ....AspSerGlyLysAspArgCysLeuHisGluSerLysGlyArgAlaHis 258
|||||
1350 GGAGCGGACAGAGCGCGCGCGCATTTGGCTGGGCGAAGACCAAGGGCAGACGCC 1399
|||||
258 isProGlnValAspProLysLeuLeuAspLysLeuHisGlnTyrPheHis 274
|||||
1400 ATCTTCATATCGACCGCGGAGGGTGGTGGCGACGCTGGCGGAGATCTATACGG 1449
|||||
275 CLeuProAsnLysLysPhePheLysLeuValGly 285
|||||
1450 GCTTTAAATTAATTTCTTAAAGATGAAGGG 1482

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 31, 2001, 08:46:39 ; Search time 106.58 seconds
(without alignments)
2317.987 Million cell updates/sec

Title: US-09-557-262-3

Perfect score: 1305

Sequence: 1 cgcgcctcagtaattgaagg... gaactaaatatttactga 1305

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 94555562 residues

Word size: 16

Total number of hits satisfying chosen parameters: 149

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: Issued_Patents_NA: *

- 1: /cgn2.6/prodata/2/ina/5A_COMB.seq: *
- 2: /cgn2.6/prodata/2/ina/5A_COMB.seq: *
- 3: /cgn2.6/prodata/2/ina/6A_COMB.seq: *
- 4: /cgn2.6/prodata/2/ina/6B_COMB.seq: *
- 5: /cgn2.6/prodata/2/ina/PTUS_COMB.seq: *
- 6: /cgn2.6/prodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	19	1.5	558	4	US-08-998-416-906	Sequence 906, App
2	18	1.4	1451	4	US-09-276-531-28	Sequence 28, Appl
3	18	1.4	7812	4	US-09-368-590-1	Sequence 1, Appl
4	18	1.4	10747	2	US-08-147-777-1	Sequence 1, Appl
5	18	1.4	10747	3	US-08-452-872-1	Sequence 1, Appl
6	18	1.4	10747	5	PCT-US93-03985-1	Sequence 1, Appl
7	18	1.4	24979	2	US-08-147-777-3	Sequence 3, Appl
8	18	1.4	24979	3	US-08-452-872-3	Sequence 3, Appl
9	18	1.4	24979	5	PCT-US93-03985-3	Sequence 3, Appl
10	17	1.3	36	1	US-08-311-760A-36	Sequence 96, Appl
11	17	1.3	36	1	US-08-311-760A-271	Sequence 271, App
12	17	1.3	36	1	US-08-311-760A-272	Sequence 272, App
13	17	1.3	36	1	US-08-311-760A-273	Sequence 273, App
14	17	1.3	36	2	US-08-774-310-96	Sequence 273, App
15	17	1.3	36	2	US-08-774-310-271	Sequence 96, Appl
16	17	1.3	36	2	US-08-774-310-272	Sequence 271, App
17	17	1.3	36	2	US-08-774-310-273	Sequence 272, App
18	17	1.3	49	1	US-07-885-689A-26	Sequence 273, App
19	17	1.3	51	1	US-07-885-689A-13	Sequence 26, Appl
20	17	1.3	387	1	US-08-458-084-8	Sequence 13, Appl
21	17	1.3	387	1	US-08-205-508-8	Sequence 8, Appl
22	17	1.3	387	2	US-08-482-148-5	Sequence 5, Appl
23	17	1.3	387	5	PCT-US95-02944-5	Sequence 5, Appl
24	17	1.3	387	5	PCT-US95-02945-8	Sequence 8, Appl
25	17	1.3	576	1	US-07-885-689A-28	Sequence 28, Appl
26	17	1.3	576	1	US-08-093-383-4	Sequence 4, Appl
27	17	1.3	579	1	US-07-885-689A-36	Sequence 36, Appl

Sequence 27, Appl
Patent No. 5514646
Sequence 6, Appl
Sequence 3, Appl
Sequence 7, Appl
Sequence 12, Appl
Sequence 880, Appl
Sequence 55, Appl
Sequence 57, Appl
Sequence 24, Appl
Sequence 25, Appl
Sequence 10, Appl
Sequence 88, Appl
Sequence 139, Appl
Sequence 85, Appl

US-08-998-416-906
Sequence 906, Application Us/08948416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippseu, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Rebschunq, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998.416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/POCKET NUMBER: PE/5-30366/A/00-1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 906:
SEQUENCE CHARACTERISTICS:
LENGTH: 558 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1565Up

ALIGNMENTS



US-08-998-416-906

Query Match 1.5%; Score 19; DB 4; Length 558;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1112 ttttaaaagcatttaagg 1130
Db 295 TTTAAAAAGCATTTAAGG 313

RESULT 2
US-09-276-531-28/C
Sequence 28, Application US/0942/16531
Patent No. 6183968
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Yue, Henry
APPLICANT: Reddy, Roopa
APPLICANT: Guegier, Karl J.
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING
TITLE OF INVENTION: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/276,531
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/079,677
FILING DATE: March 27, 1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Lynn E. Murty, Ph.D.
REGISTRATION NUMBER: 42,918
REFERENCE/DOCKET NUMBER: PA-0008 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1451 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: HNT2GT01
CLONE: 491493
US-09-276-531-28

Query Match 1.4%; Score 18; DB 4; Length 1451;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 915 tacatgagtcacaaagccc 932
TACATGAGTCACAAAGCCC 279

RESULT 3
US-09-368-590-1
Sequence 1, Application US/09368590
Patent No. 6187563
GENERAL INFORMATION:
APPLICANT: Solimena, Michele
TITLE OF INVENTION: INTERACTING POLYPEPTIDES FOR
TITLE OF INVENTION: AUTOANTIGENS OF AUTOIMMUNE DISEASES
FILE REFERENCE: 101918-200 (OCR-941)
CURRENT APPLICATION NUMBER: US/09/368,590
CURRENT FILING DATE: 1998-08-04
EARLIER APPLICATION NUMBER: 60/2395,657
EARLIER FILING DATE: 1998-08-07
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO: 1
LENGTH: 7812
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(6879)
NAME/KEY: unsure
LOCATION: (100) (102)
NAME/KEY: unsure
LOCATION: (1021)...(1023)
NAME/KEY: unsure
LOCATION: (2266)...(2268)
US-09-368-590-1

Query Match 1.4%; Score 18; DB 4; Length 7812;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 138 gcgcggtgctgctgtgg 155
Db 2045 gcgcggtgctgctgtgg 2062

RESULT 4
US-08-147-777-1
Sequence 1, Application US/08147777
Patent No. 5914265
GENERAL INFORMATION:
APPLICANT: Roop, Dennis R.
APPLICANT: Rothnagel, Joseph A.
APPLICANT: Greenhalgh, David A.
APPLICANT: Yuspa, Stuart H.
TITLE OF INVENTION: KRPATIN V1 EXPRESSION VECTORS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: LYON & LYON
STREET: 611 West Sixth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90017
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: IBM MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/147,777
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below:


```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:475268"
/clone_lib="Soares mouse embryo NBME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total tetus"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
was primed with a Not I oligo(dT) primer [5',
TGTTCACCACTCGATGGGAGCGCGGAAATTTTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."
136 a 134 c 114 g 124 t

```

```

Query Match          30.1%;  Score 508;  DB 1;  Length 508;
Best Local Similarity 100.0%;  Pred. No. 4.8e-257;
Matches 508;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

QY	875	gaagctggctgaacctggactcaaaagctctcaaacgcagcctgtaccatcacacatg	934
Db	1	GACGGTCGGCTCAACCTGGACTACAAAGCTCTCAACCGCAAGCTGTACCATGCACACATG	60
QY	935	ctgaacctggctgcgctttttcccgcttggcgccatccacattggatgagcaccgcctc	994
Db	61	CTGAACCTGGCTCGCTTTTCCCGCTTGGCCACATCCACATTCTGGATGCGACCGCCTC	120
QY	995	atcagagaccctttccctggagatccagaagctggaagattctgaagctttctccacag	1054
Db	121	ATCAGAGACCCCTTTCCTTCGATCCCAAGGTCGAAGATTCCTGAAGCTTTCTCCACAG	180
QY	1055	atcaacgctggaactctactttaacaaaccaaggctctctactgcctgcggacagt	1114
Db	181	ATCAACGGCTCGAACTTCTACTTTAAACAAACCAAGGGTCTTACTTGGCTGGGACAGT	240
QY	1115	ggcaaggaccgctgctttacacgagctccaaaggccggcgccacccccagctggaatccaaa	1174
Db	241	GGCAAGGACCGCTGCTTACACGAGCTCCCAAGGCCGGCCGCACCCCGCAGTGGATCCCAA	300
QY	1175	ctacttgataaactgcagaaactcttctatgagcccaataagaatattttcaagctctgt	1234
Db	301	CTACTTGTATAACTCGACGAATCTTCTATGAGCCAAATGAAGAATTTTCAAGTCGTG	360
QY	1235	ggcagaacattcgactggcactgaattgcgctctctagctgcggaaatttccgtgtgt	1294
Db	361	GGCAGAACTTCGACTGGCACTGGCACTGATTTGGCGCTCTCTAGCGCTCGGAGCTTTCTGTTGT	420
QY	1295	taactctggtgtacattctgaagggggaggaataataattttaaaaggcatttaagcta	1354
Db	421	TAACTCTGGTGTACATCTGAAGGGGGAGAGAAATTAATTTTAAAGAGCATTTAAGCTA	480
QY	1355	taatttatgtataaaccccaaatgac	1382
Db	481	TAAATTTATTGTAAACCCCAAAATGAC	508
RESULT	4		
LOCUS	W62484		
DEFINITION	md72cl2.r1 Soares mouse embryo N0Me13.5 14.5 Mus musculus cDNA clone IMAGE:373942 5' similar to PIR:A49733 A49733		
FEATURES	glycosaminoglycan N-acetylglucosaminyl N-deacetylase ;, mRNA sequence.		
ACCESSION	W62484		

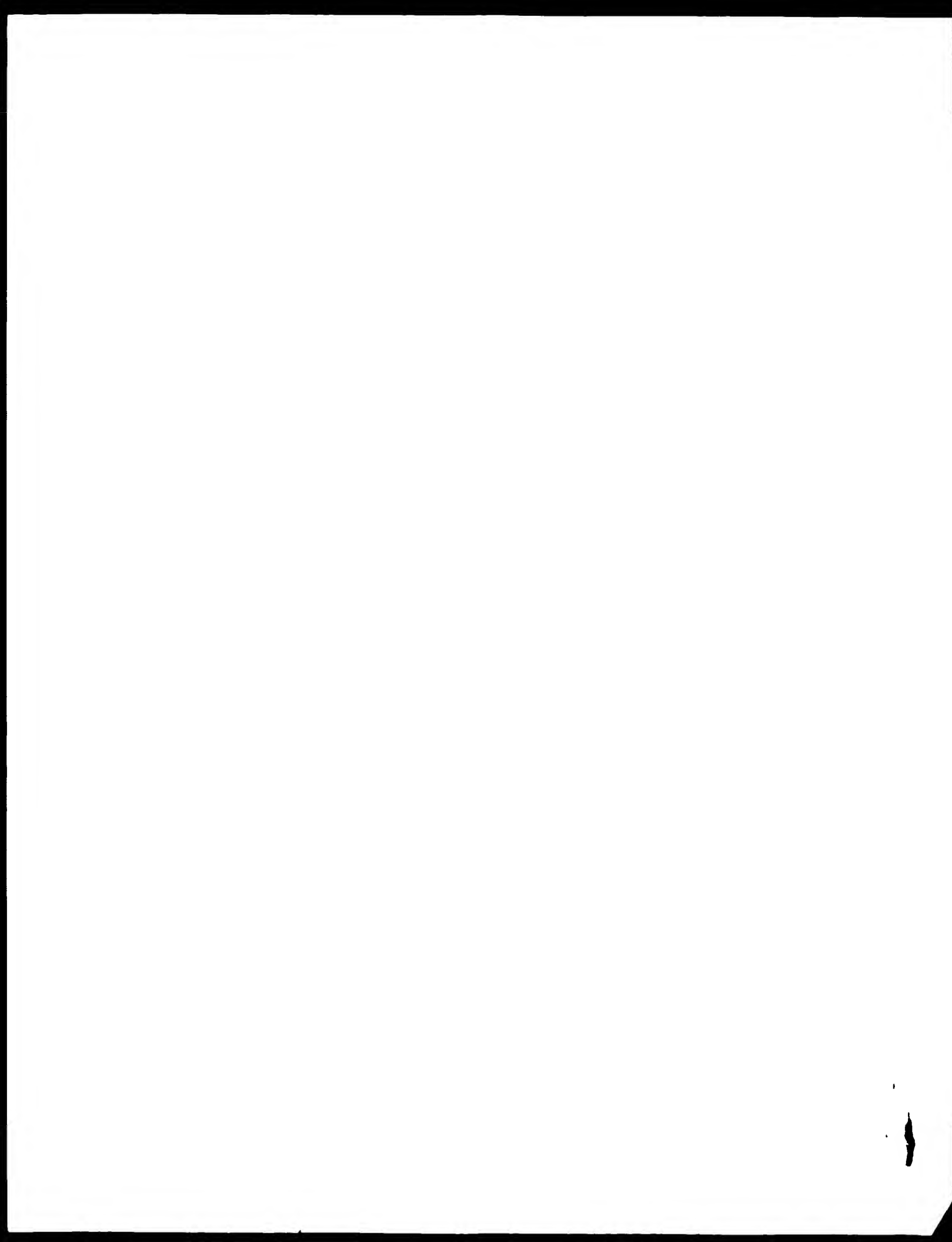
```

VERSION      W62484.1  GI:1369243
KEYWORDS     EST.
SOURCE       house mouse.
ORGANISM     Mus musculus

REFERENCE    Fukuyuta, T. et al. (1990)
AUTHORS      Marmè, D., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dunaway, T.,
              Geisler, S., Kralova, T., Lacy, M., Le, M., Martin, J., Morris, M.,
              Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
              Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
              Waterston, R.
TITLE        The WashU-HMI Mouse EST Project
JOURNAL      Unpublished (1996)
COMMENT      Contact: Maria M/Mouse EST Project
              WashU-HMI Mouse EST Project
              Washington University School of Medicine
              444 Forest Park Parkway, Box 8501, St. Louis, MO 63109
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: mouseest@watson.wustl.edu
              This clone is available royalty-free through LNL; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              MGI:235374
Seq primer:  ETPrimer
High quality sequence stop: 344.
Location/Qualifiers
1..490
/organism="Mus musculus"
/strain="c57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:373942"
/clone_lib="Soares mouse embryo NMEL1.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13 e-14 5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pMTT30-Pac (Pharmacia) with a modified
polylinker; Site1: Not I; Site2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'-
TGTTCACCAATCGAAGTGGAGGCGGCGGAAATTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos (total RNA provided by Minora Ko, Wayne
State Univ., from 2); double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Rpl I and Eco RI sites of the modified
pMT3 vector. Library went through one round of
normalization, and was constructed by Pedro Soares and
M.Fátima Bonaldo."
127 a 129 c 122 y 122 t

```

Query Match	29.1%	S.N. 4501	12 150;	Length 490;
Best Local Similarity	100.0%;	Pred. No. 1.6e-247;		
Matches 490;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
875	gaaggttggtgagcctgagctacagagctctcaagcgcagcctgtacatgacacatg	934		
1	GACGGTGGCGTGAACCTGGACCTACAAAGGCTCTCAACCGCAGCCTGTACCATGTGACATG	59		
935	CTgaactggcgtcggttttttccgttgggcgcacatccacatttgatagcgaagccttc	994		
61	CTGAACCTGGCGTGGCTTTTTCGGCTTGGCGGCACATCCATCCGATGTGCATGGAGATGGCTT	120		
995	atcagagagacctttccctcgagatccagaaaggtccgaagatctctgaagctttctctacacg	1054		
121	ATTAAGATACGCGTTTCGGCTGAGATCAAGAGAGATTCGAAAGATTCCTTCTCTACAG	180		
1055	atcaagagctctgaactttactcttttaacaaagcgaagggctctactgctctgaggaacat	1114		
181	ATCAAGCGCTCGAAGCTTCTACTTTTAAACAAACCAAGGGCTCTTATCTGCTCTGGGACAGT	240		



1115 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 1174
 241 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 300
 1175 ctacttaataaactgcacgaatactilcatagcagcaaaataagaaatctttcaagctcgtg 1234
 301 ctacttataaactgcacgaatactilcatagcagcaaaataagaaatctttcaagctcgtg 360
 1235 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 1294
 361 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 420
 1295 taactcttggtgacatctgaaggggggagaaataaatttaaaagcatttaagcta 1354
 421 taactcttggtgacatctgaaggggggagaaataaatttaaaagcatttaagcta 480
 1355 taattattt 1364
 481 taattattt 490

RESULT 5
 1852449/c
 OCUS
 DEFINITION UI-M-BH0-ajo-h-09-0-UI s1 NIH_BMAP_M.S1 Mus musculus cDNA clone
 UI-M-BH0-ajo-h-09-0-UI 3', mRNA sequence.
 CCESSION A1852449.1 GI:5496320
 ERSION A1852449.1
 EYWORDS EST
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 482)
 Bonaldo,M.F., Lennon,G. and Soares,M.R.
 Normalization and subtraction: Two approaches to facilitate gene
 discovery
 Genome Res 6 (9), 791-806 (1996)
 9704477
 Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7150, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mestr@mail.nih.gov
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to identify it as a clone from the
 normalized hippocampus library cDNA library Preparation: M.B.
 Soares lab clone distribution: NIH_BMAP cDNA clones will be made
 available by the means that is soon to be determined. When NIH
 determines the means for distribution of the BMAP cDNA clones, this
 record will be updated accordingly when that means is determined.
 Seq primer: M13 Forward
 POLA=yes
 FEATURES
 source Location/Qualifiers
 1. .482
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-BH0-ajo-h-09-0-UI"
 /clone_1lb="NIH_BMAP_M.S1"
 /env_stage="27-32 days"
 /lab_host="pH108 (Life Technologies)"
 /note="Vector: pT73D-pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The
 NIH_BMAP_M.S1 library is a subtracted library derived from
 a mixture of normalized libraries from ten regions of the
 mouse brain (cerebellum, brain stems, olfactory bulbs,
 hypothalamus, cortex, amygdala, basal ganglia, pineal

gland, striatum, hippocampus). The driver used for
 subtraction consisted of a pool of 20,000 cDNA clones
 obtained from non-normalized and normalized libraries of
 these ten regions of the mouse brain.
 TAG_L1B-NIH_BMAP_M.S1
 TAG_TISSUE=hippocampus
 TAG_SEQ=ITCGA*
 BASE COUNT 166 a 80 g 147 t
 ORIGIN
 Query Match 27.8%; Score 469; DB 102; Length 482;
 Best Local Similarity 100.0%; Pred. No. 2e-236;
 Matches 469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1145 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 1204
 DB 482 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 423
 QY 1205 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 1264
 DB 422 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 363
 QY 1265 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 1324
 DB 362 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 303
 QY 1325 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 1384
 DB 302 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 243
 QY 1385 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 1444
 DB 242 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 183
 QY 1445 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 1504
 DB 182 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 123
 QY 1505 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 1564
 DB 122 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 63
 QY 1565 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 1613
 DB 62 ggcaggaacccgctgcttacagcagtcacaaagcggcgccgaccccccaggtggatccaaa 14

RESULT 6
 BG063038/c
 LOCUS BG063038.3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
 DEFINITION H3001502.3', mRNA sequence.
 ACCESSION BG063038
 VERSION BG063038.1 GI:12545524
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 512)
 Karqui,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
 J.S., Carter,M.G. and Ko M S H
 Verification and initial annotation of NIA mouse 15K cDNA clone set
 Unpublished (2001)
 Other ESTs: H3001502-5
 Contact: George J. Karqui
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@lgsun.grc.nia.nih.gov
 This clone set has been freely distributed to the community. Please
 visit <http://lgsun.grc.nia.nih.gov/cDNA/15k.html> for details.



STIC-Biotech/ChemLib

From:
Sent:
To:
Subject:

Shukla, Ram
Monday, August 27, 2001 1:14 PM
STIC-Biotech/ChemLib
09/557,262

Please search for:

nucleic acid of Seq ID NO 1
nucleic acid encoding SEQ ID NO 2
nucleic acid of Seq ID NO 3
nucleic acid encoding SEQ ID NO 4
Nucleic acid encoding residues 21-52, 260-269, 250-276, 53-311 or 21-307 of SEQ ID NO 2
Nucleic acid encoding residues 21-48, 256-265, 246-272, 49-307 or 21-303 of SEQ ID NO 2
Oligosearch (16 mer) for SEQ ID NO 1 and 3

Thanks.

Patent Examiner
AU 1632
Crystal Mail 1, Rm 12E03
USPTO
Phone: (703) 305-1677
Fax: (703) 746-3103

10E12
m3

Edward Grant
Technical Info Specialist
STIC / Biotech
CMI 12C14 Tel: 305-9203

49702



$$p_{\tau} =$$

FILE, MEDLINE, ENTERED AT 17:39:16 ON 26 OCT 2001

183622	2	EFFICIENT	1
	3	IN ANT 13 AND 14 AND 17	2
	4	3-05T	3
	5	19 NOV EX-1996	4
	6	19 NOV EX-1997	5
	7	HEPARIN SULPHATE PROTEOGLYCAN OR HEPARIN PROTEOGLYC?	6
	8	13 ANT 17	7
	9	21 NOV EX-1997	8
	10	21 NOV EX-1996	9
	11	21 NOV EX-1997	10
	12	21 NOV EX-1996	11
	13	21 NOV EX-1997	12
	14	21 NOV EX-1996	13
	15	21 NOV EX-1997	14
	16	21 NOV EX-1996	15
	17	21 NOV EX-1997	16
	18	21 NOV EX-1996	17
	19	21 NOV EX-1997	18
	20	21 NOV EX-1996	19
	21	21 NOV EX-1997	20
	22	21 NOV EX-1996	21
	23	21 NOV EX-1997	22
	24	21 NOV EX-1996	23
	25	21 NOV EX-1997	24
	26	21 NOV EX-1996	25
	27	21 NOV EX-1997	26
	28	21 NOV EX-1996	27
	29	21 NOV EX-1997	28
	30	21 NOV EX-1996	29
	31	21 NOV EX-1997	30
	32	21 NOV EX-1996	31
	33	21 NOV EX-1997	32
	34	21 NOV EX-1996	33
	35	21 NOV EX-1997	34
	36	21 NOV EX-1996	35
	37	21 NOV EX-1997	36
	38	21 NOV EX-1996	37
	39	21 NOV EX-1997	38
	40	21 NOV EX-1996	39
	41	21 NOV EX-1997	40
	42	21 NOV EX-1996	41
	43	21 NOV EX-1997	42
	44	21 NOV EX-1996	43
	45	21 NOV EX-1997	44
	46	21 NOV EX-1996	45
	47	21 NOV EX-1997	46
	48	21 NOV EX-1996	47
	49	21 NOV EX-1997	48
	50	21 NOV EX-1996	49
	51	21 NOV EX-1997	50
	52	21 NOV EX-1996	51
	53	21 NOV EX-1997	52
	54	21 NOV EX-1996	53
	55	21 NOV EX-1997	54
	56	21 NOV EX-1996	55
	57	21 NOV EX-1997	56
	58	21 NOV EX-1996	57
	59	21 NOV EX-1997	58
	60	21 NOV EX-1996	59
	61	21 NOV EX-1997	60
	62	21 NOV EX-1996	61
	63	21 NOV EX-1997	62
	64	21 NOV EX-1996	63
	65	21 NOV EX-1997	64
	66	21 NOV EX-1996	65
	67	21 NOV EX-1997	66
	68	21 NOV EX-1996	67
	69	21 NOV EX-1997	68
	70	21 NOV EX-1996	69
	71	21 NOV EX-1997	70
	72	21 NOV EX-1996	71
	73	21 NOV EX-1997	72
	74	21 NOV EX-1996	73
	75	21 NOV EX-1997	74
	76	21 NOV EX-1996	75
	77	21 NOV EX-1997	76
	78	21 NOV EX-1996	77
	79	21 NOV EX-1997	78
	80	21 NOV EX-1996	79
	81	21 NOV EX-1997	80
	82	21 NOV EX-1996	81
	83	21 NOV EX-1997	82
	84	21 NOV EX-1996	83
	85	21 NOV EX-1997	84
	86	21 NOV EX-1996	85
	87	21 NOV EX-1997	86
	88	21 NOV EX-1996	87
	89	21 NOV EX-1997	88
	90	21 NOV EX-1996	89
	91	21 NOV EX-1997	90
	92	21 NOV EX-1996	91
	93	21 NOV EX-1997	92
	94	21 NOV EX-1996	93
	95	21 NOV EX-1997	94
	96	21 NOV EX-1996	95
	97		

FILE 'USPABEUTL', ENTERED AT 17:57:41 ON 06 OCT 2001

L26	11 S TIG
L27	1306 E HID
L28	0 S (NUCLEIC ACID OR DNA OR RNA)
L29	7 S (NUCLEIC ACID OR DNA OR RNA) AND L26



L Number	Hits	Search Text	DB	USPAT:	US-PGPUB:	EPO: JPO:	DERWENT	Time stamp
1	6	3-ost or o-sulfoltransferase	0	(3-ost or o-sulfoltransferase) same nucleic ADJ acid same host same vector	USPAT:	US-PGPUB:	DERWENT	2001/10/26 18 08
7	0	(3-ost or o-sulfoltransferase) same nucleic ADJ acid same	0	(3-ost or o-sulfoltransferase) same nucleic ADJ acid same	USPAT:	US-PGPUB:	DERWENT	2001/10/26 18 09
19	0	(3-ost or o-sulfoltransferase) same nucleic ADJ acid same vector	0	(3-ost or o-sulfoltransferase) same nucleic ADJ acid same	USPAT:	US-PGPUB:	DERWENT	2001/10/26 18 09
25	0	(3-ost or o-sulfoltransferase) same nucleic ADJ acid same host	0	(3-ost or o-sulfoltransferase) same nucleic ADJ acid same	USPAT:	US-PGPUB:	DERWENT	2001/10/26 18 09
13	2	(3-ost or o-sulfoltransferase) same nucleic ADJ acid	2	(3-ost or o-sulfoltransferase) same nucleic ADJ acid	USPAT:	US-PGPUB:	DERWENT	2001/10/26 18 10
31	2	transgen\$ and (3-ost or o-sulfoltransferase)	2	transgen\$ and (3-ost or o-sulfoltransferase)	USPAT:	US-PGPUB:	DERWENT	2001/10/26 18 10



L	Number	Hits	Search Text	DB	Time stamp
7		2	(3-ost or o-sulfotransferase) same nucleic ADJ acid	USPAT; US-PGPUB;	2002/07/25 14:51
13		154108	(3-ost or o-sulfotransferase) and human or mouse	USPAT; US-PGPUB;	2002/07/25 14:50
19		154109	3-ost or o-sulfotransferase and human or mouse	USPAT; US-PGPUB;	2002/07/25 14:51
25		6	transgen\$ and (3-ost or o-sulfotransferase)	USPAT; DERWENT EPO; JPO; US-PGPUB;	2002/07/25 14:51
	1	10	3-ost or o-sulfotransferase	USPAT; DERWENT EPO; JPO; US-PGPUB;	2002/07/25 14:51
		6	3-ost or o-sulfotransferase	USPAT; DERWENT EPO; JPO; US-PGPUB;	2002/07/25 14:49
		0	(3-ost or o-sulfotransferase) same nucleic ADJ acid same host same vector	USPAT; DERWENT EPO; JPO; US-PGPUB;	2001/10/26 18:09
		0	(3-ost or o-sulfotransferase) same nucleic ADJ acid same vector	USPAT; DERWENT EPO; JPO; US-PGPUB;	2001/10/26 18:09
		0	(3-ost or o-sulfotransferase) same nucleic ADJ acid same host	USPAT; DERWENT EPO; JPO; US-PGPUB;	2001/10/26 18:09
		2	(3-ost or o-sulfotransferase) same nucleic ADJ acid	USPAT; DERWENT EPO; JPO; US-PGPUB;	2002/07/25 14:50
		2	transgen\$ and (3-ost or o-sulfotransferase)	USPAT; DERWENT EPO; JPO; US-PGPUB;	2002/07/25 14:51
	22101	signal ADJ sequence		USPAT; US-PGPUB;	2001/10/30 14:41
	3	("5541095").PN.		USPAT; DERWENT EPO; JPO; US-PGPUB;	2001/10/30 14:41
	2	("5817487").PN.		USPAT; DERWENT EPO; JPO; US-PGPUB;	2001/10/30 14:42
	0	("11 same 17").PN.		USPAT; US-PGPUB;	2001/10/30 14:42
	-	DERWENT		DERWENT	



STIC-Biotech/Chemlib

From: Shukla, Ram
Sent: Monday, July 15, 2002 1:39 PM
To: STIC-Biotech/Chemlib
Subject: 09/557,262

Please search for:
NA encoding 21-307 of SEQ ID NO 2
NA encoding 53-311 of SEQ ID NO 2
NA encoding 250-276 of SEQ ID NO 2
NA encoding 21-303 of SEQ ID NO 4
NA encoding 49-307 of SEQ ID NO 4
NA encoding 246-272 of SEQ ID NO 4.

Thanks

Patent Examiner
AU 1632
Crystal Mail 1, Rm 12E03
Mailbox 12E12
USPTO
Phone: (703) 305-1677
Fax : (703) 746-3103

POINT OF CONTACT:
PAUL SCHULWITZ
TECHNICAL INFO. SPECIALIST
CM1 6806 TEL. (703) 305-1954

TYPE OF SEARCH:

NA Sequences: 6
AA Sequences:
Structures:
Bibliographic:
Litigation:
Full text:
Patent Family:
Other:

Searcher:
Phone:
Location:
Date Picked Up:
Date Completed:
Searcher Prep/Review:
Clerical:
Online time:

VENDOR/COST (where applic.)
STN:
DIALOG:
Questel/Orbit:
DRLink:
Lexis/Nexis:
Sequence Sys.:
WWW/Internet:
Other (specify):

70881

